Modeling Emotional Blends from an Appraisal Theory Perspective

By

Weiqiang Qian

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 13, 2021

Nashville, Tennessee

Approved:

Craig A. Smith, Ph.D.

Leslie D. Kirby, Ph.D.

Duane G. Waston, Ph.D.

Daniel T. Levin, Ph.D.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Theorists have approached the study of emotion with the perspective that various emotions serve distinctive adaptive functions to benefit humans (Smith & Lazarus, 1990; Gross & Barret, 2011). Emotions improve personal well-being by directing people to address environmental challenges. For example, anger prompts people to address harms caused by others through removing the sources of harm, whereas guilt directs people to deal with damages brought by themselves by righting their own wrongs (Smith, 1991). Researchers have long studied emotions by picking an emotion and investigating properties of that specific emotion. However, when an emotion is present, rarely does it occur completely isolated from other emotions. Hence, theoretical work that focuses on singular emotions runs the risk of not taking into account the broader context of emotional experience, thereby limiting the generalizability of the resulting theory. The blending of multiple emotions, called emotional blends in some literatures, has been documented to be rather prevalent in emotional experience (Smith & Ellsworth, 1987; Ellsworth & Smith, 1988a, 1988b; Zelenski & Larsen, 2000; Robinson & Clore, 2001).

Currently, the amount of research on emotional blending is incommensurate with the prevalence of the phenomenon. Emotional blends are largely unstudied beyond people reporting that they exist (Berrios, Totterdell, & Kellett, 2015). So far, there has not been a systematic treatment of emotional blends. Much is unknown about emotional blends, such as what emotions tend to co-occur, or what adaptive functions the co-occurring emotions serve. Although a special case of emotional blends, consisting of polar opposite emotions like anger and happiness, has gained some traction in recent years (Oatley & Johnson-Laird, 1996; Russell & Caroll, 1999; Russell, 2003, 2017; Larsen & McGraw, 2014; Berrios, Totterdell, & Kellett, 2015), most of the emotional blends involving emotions of similar valence have not been documented. I believe to truly understand emotional blends, it is necessary to document what emotional blends people generally experience and provide a theoretical account that explains special cases and general cases alike.

A good theoretical framework to explain emotional blends is appraisal theory. Under appraisal theory, what elicits emotional responses is the individual's interpretation of the personal implications of the environment, rather than the objective environment itself. Various dimensions of appraisal capture different implications from human-environmental relations for subjective well-being, which in turn elicit emotions to address the implications (Smith & Lazarus, 1990). Anger arises when environmental challenge is unpalatable and other-induced, whereas guilt happens when unpleasantness is self-induced. Appraisal theory has long been studied systematically for each of the emotions individually. Ample theoretical reviews have portrayed appraisal foundations for lists of emotions in their singular forms (Roseman, Spindel, & Jose, 1990; Smith, 1991). Table 1, produced from Smith (1991), maps the key appraisals associated with 4 negative and 5 positive emotions. Table 1 also contains the summary of environmental implication each emotion addresses, which is conceptually equivalent to the core relational theme described by Smith (1991).

The appraisals used to explain the emotions in Table 1 follow the Smith and Lazarus view of appraisal theory, which contains 7 dimensions of appraisals. Relevance appraisal taps into whether the environmental demands at hand matter to personal well-being, while congruence appraisal characterizes the degree to which environmental implications align with personal goals to promote well-being. Environmental stimuli completely irrelevant to personal well-being should not elicit emotion, nor should stimuli that carry no positive or negative implications for well-being. These two appraisals are fundamental to the adaptive value of emotions, without which there would be no environmental concerns to tend to. Once a situation poses an opportunity to formulate emotions, appraisal components other than relevance and congruence capture coping information from the environmental stimuli to differentiate specific emotions and allow for coping behaviors that directly address the environmental concerns. These appraisals include self-accountability, other-accountability, problem-focused coping potential (PFCP), accommodative-focused coping potential (AFCP), and future expectancy. The two accountability appraisals describe whether the self or someone else is responsible for creating the situational concerns. They help to clarify who is to blame or worthy of praise. The source of blame can differ from the person who carried out the action, as harm done unintentionally lessens the blame on the person (Ames & Fiske, 2013). PFCP provides a self-assessment over the person's capability of solving environmental challenges, while AFCP taps into a given individual's evaluation of their own ability to adjust to the situation, no matter what happens to the situation. The two different coping potentials provide a coverage for the possible behavioral options that include modifying the situation to conform to own desire, or modifying oneself to conform to the situational development. Future expectation describes expected outlook of the situation.

Table 1. Major emotions and their associated appraisals according to Smith (1993)

Emotion	Summary Environmental Implication	Appraisals
Anger	Other-harm	High relevance Low congruence Other-accountability
Guilt	Self-harm	High relevance Low congruence Self-accountability
Anxiety	Threat	High relevance Low congruence Low accommodative-focused coping potential
Sadness	Loss	High relevance Low congruence Low problem-focused coping potential Low future expectancy
Hope	Potential improvement	High relevance Low congruence High future expectancy
Challenge	Active engagement	High relevance Low congruence High problem-focused coping potential High future expectancy
Happiness	Success	High relevance High congruence
Pride	Self-benefit	High relevance High congruence Self-accountability
Gratitude	Other-benefit	High relevance High congruence Other-accountability

In a similar fashion to Table 1, an emotional blend adaptation of appraisal theory can help to map major emotional blends onto their characteristic environmental implications. Because appraisal theory models the environmental implications of a specific emotional state, for it to be used to model emotional blends, it requires establishing that each emotional blend contains a simultaneous experience of multiple emotions. Although there are arguments for different emotions switching rapidly to allow for only one emotion at any given time (Russell, 2017), evidence and argument for the simultaneous experience of multiple emotions as a coherent emotional state have grown strong over time. Experimental inductions of opposing-valence emotions with continuous and indivisible stimuli have successfully elevated subjective reporting of concurrent emotions (Hunter, Schellenberg, & Schimmack, 2008; Larsen & McGraw, 2011). On a subjective experiential level, induction of opposing valence emotions has shown a rapid succession of emotions is subjectively indistinguishable from a parallel presence of multiple emotions (Schimmack & Colcombe, 2007). Theory and evidence from a neuroscience perspective have given further biological support to a distributed processing structure for valence that allows for parallel valence states (LeDoux, 1991; Ohman, 1999; Man, Nohlen, Melo, & Cunningham, 2017). The support for the coherence in the emotional experience of multiple emotions paves the way for treating emotional blends as unitary experiential states featuring simultaneous emotions.

An adaptation to the appraisal theory to account for emotional blends will then look somewhat similar to how appraisals are organized around each singular emotion: emotional blends are tied to specific profiles of appraisals that differentiate one blend from another. While previous research have supported a coherent emotional state for emotional blends, nothing is known so far about how appraisals relevant to each emotion in a blend are synthesized to form the appraisal profile for the blend. On one hand, appraisals related to each emotion in a blend could be retained to form the overall appraisal profile without modification. This would mean that each emotion still functions to serve the adaptive function as it would within singular emotion settings, except now due to the co-occurrence of emotions, different sets of appraisals are merged into one. I call this way of blending emotions the additive model. Some mechanisms for the additivity include when emotions with conflicting appraisals are blended, the conflicting appraisals become averaged, which leads to the averaged appraisal still associated in the same way with the emotions in the blend as it would with the singular emotion. On the other hand, when emotions are blended in a non-additive fashion, the underlying appraisal structure associated with each emotion might change to accommodate for the integration of appraisal from different emotions. Previous studies have already shown that it is possible to have some variability in appraisal structure of singular emotions (Kuppens, Van Mechelen, Smits, & De Boeck, 2003; Kuppens, Van Mechelen, Smits, De Boeck, & Ceulemans, 2007), paving ways for enabling appraisal structure of a blended emotion to change when other emotions are present. With changes to the supporting appraisals, singular emotion in a blend may serve its adaptive function differently from what it does within singular emotion settings, allowing for the blend to address particular environmental implications not served by each of the constituting emotions on its own. Both the additive and non-additive ways of blending emotions result in a single appraisal profile supporting the experience of multiple emotions. The common appraisal profiles are unique entities that address specific environmental implications, although the uniqueness have different meanings for the additive and non-additive blending model. Where in both models a specific appraisal profiles addressing a valid combination of environmental implications is uniquely linked to a particular configuration of emotions that can be realistically experienced, the non-additive model further adds uniqueness to how appraisals are linked to emotions within the particular emotion elicitation context, resulting in a summation of emotional experience that is more than its parts.

Table 2. Proposed major emotional blends, environmental implications, and associated appraisals

Blend	Emotions	Environmental Implications	Appraisals
Self-negativity blend	Guilt, anger (self-directed), sadness	Self-harm, loss	High relevance Low congruence High incongruence High self-accountability Low future expectancy Low problem-focused coping
Dual negativity blend	Guilt, anger (self-directed), anger (other-directed), sadness, anxiety, hope	Self-harm, other-harm, loss, threat, potential improvement	High relevance Low congruence High incongruence High self-accountability High other-accountability Medium future expectancy Low problem-focused coping
Anger blend	Anger (other-directed), sadness, anxiety	Other-harm, loss, threat	High relevance Low congruence High incongruence High other-accountability Low future expectancy Low problem-focused coping Low accommodative-focused coping
Anxiety blend	Anxiety, sadness, challenge	Threat, loss, active engagement	High relevance Low congruence High incongruence Medium future expectancy Medium problem-focused coping Low accommodative-focused coping
Mild positivity blend	Calm/Tranquility, mild other positive emotions	Lack of threat, openness to opportunity	Low relevance Medium-to-high congruence Low incongruence High future expectancy.
Opportunity blend	Challenge, hope, anxiety	Active engagement, potential improvement, threat	High relevance Low congruence High incongruence High future expectancy High problem-focused coping Low accommodative-focused coping
Self-benefit blend	Happiness, pride	Success, self-benefit	High relevance High congruence Low incongruence High self-accountability
Other-benefit blend	Happiness, gratitude	Success, other-benefit	High relevance High congruence Low incongruence High other-accountability
Bittersweet blend	Happiness, sadness	Success, loss	Medium relevance High congruence High incongruence Low future expectancy Low problem-focused coping

Drawing information and inspiration from previous research on the coexistence among emotions (Ellsworth & Smith, 1988a; Ellsworth & Smith, 1988b; Robinson & Clore, 2001), I propose a few theoretically predicted emotional blends and their associated appraisals in Table 2. Replacing a list of singular emotions is the list of emotional blends that should be featured prominently in the emotional life of people. In each predicted blend, the appraisals capture a complex set of implications within the emotion eliciting context that exceeds what a singular emotion can reflect. The speculated environmental implications for each predicted blend is listed within Table 2, alongside the predicted appraisals based on singular emotion appraisal-emotion relations from Table 1, and emotions predicted to follow the appraisals.

The list features three categories of emotional blends. The first category includes 4 negative emotional blends. Although the descriptor for some blends centers on a particular singular negative emotion, each blend contains multiple emotions that together describe a person's emotional state. The boundaries among different emotional blends are hardly clear-cut: negative emotional blends can have many overlapping negative emotions, nor are positive emotions completely excluded from negative emotional blends. The first predicted blend, the self-negativity blend, contains negative emotions directed at oneself including guilt, self-directed anger, and sadness (Ellsworth & Tong, 2006). Situational implications that lead to the self-negativity blend would include harm brought by oneself, along with a sense of loss. The dual negativity blend is similar to the self-negativity blend but differs in that people also want to attribute present obstacles to other social agents or uncontrollable environmental influences in addition to themselves (Smith & Ellsworth, 1987). Beyond the emotions from the self-negativity blend, the dual negativity blend also has other-directed anger that arises from harm brought by others, anxiety that arises from a perceived sense of threat, as well as hope due to people perceiving failures as somewhat unfair to defining their capability while looking forward to potential improvements. The anger blend arises in a situation that leads people to perceive other-harm, in addition to a sense of loss and threat. The emotions in the blends include other-directed anger, anxiety, and sadness. The anxiety blend contains anxiety, sadness, and challenge. The type of situation that induces the anxiety blend imparts a sense of threat and loss to people, along with an active engagement to deal with the situation.

The second category includes 4 positive emotional blends. A mild positivity blend represents the normative state people operate in for the majority of time (Diener & Diener, 1996). This blend arises in situations where there is no imminent threat, while people stay open to potential opportunities for personal growth (Fredrickson, 2004). An opportunity blend involves not only positive emotions like challenge and hope as people engage actively to make improvements to the situation, but also anxiety due to perceived threat within the environment (Smith & Ellsworth, 1987; Berrios, Totterdell, & Kellett, 2015). The two other positive blends that feature happiness have different emphasis on accountability, with one involving pride and the other involving gratitude (Smith, Tong, & Ellsworth, 2014). The contextual implications for the two blends besides success, are benefits due to self versus benefits due to others. While the first two large categories of emotional blends often involve emotions that are not typically considered to align with other emotions in the blends, like anxiety in the opportunity blend and hope in the dual negativity blend, these blends nevertheless are predominantly positive or negative.

A final category resembles the emotional experiences investigated in the mixed or dialectic emotion literature, involving both positive and negative emotions within the same emotional experience. The emotional blends in this category are experienced with lower frequency than compared to other categories of blends (Russell & Carroll, 1999; Larsen, McGraw, & Cacioppo, 2001). At the moment this category only contains one emotional blend, the bittersweet blend. This blend involves happiness and sadness at the same time, mostly experienced in situations like graduations (Ersner-Hershfield, Mikels, Sullivan, & Carstensen, 2008; Larsen, McGraw, & Cacioppo, 2001). In such situations, the contextual implications include both a sense of success and a sense of loss. Because of the ambiguity in the opposing environmental implications, situations leading to bittersweet may not be particularly urgent, hence the relevance appraisal is predicted to be a medium level. This category can potentially contain many more blends. For example, feelings of both pride and pity in a situation where a rival is trounced. While such additional blends are not listed in Table 2 due to relative rarity, they nevertheless attest to the variety of emotional blends which should generate future research interest.

Major Questions and Plans for Study

Appraisal theory earns its popularity with its utility in explaining common singular emotions. To become comparable with the singular emotion version, the emotional blends extension of appraisal theory needs answers to two questions. First and arguably the biggest question would be the makeup of the major emotional blends. Unlike singular emotions, where appraisal theorists can target a few major emotions such as anger and regret to build appraisal profiles for emotions, many of which come from commonly used vocabularies, it is difficult to clearly identify the emotional blends people experience. Where for some blends like bittersweet there are prevalent words used to describe the particular blend, most blends do not have widely adopted vocabularies. As such, it is not straightforward to find anchors to emotional blend analysis through identifying words describing blends. However, an alternative to finding those anchors in the emotional blend space is to use a data-driven approach to document emotions that commonly co-occur in the realistic emotional experience reported by people. As previously discussed for creating the emotional blend appraisal theory, a unique appraisal pattern integrates various aspects of environmental implications to induce a specific emotional blend, which in turn serves useful function to address the multiple environmental implications. The data-driven approach works on this utility-driven assumption: realistic emotional experience reported by people must have served some utility to people in their environment, hence whatever pattern of co-occurring emotions consistently observed from a sample of realistic emotional experience will be capturing the utility of that particular co-occurrence.

The data-driven approach to work with observational data offers a first look at what are some of the emotional blends that are predicted, or not predicted, by the theory-driven approach seen in Table 2. The match between observation and theory would establish a foundation of an emotional blend system with distinct entries that are seen widely in observational data. This foundation in turn offers a clear target to seek appraisals associated with each of the prevalent emotional blends. The data-driven identification of emotional blends may counter what has been predicted theoretically. Under such circumstances, the data-driven approach will provide feedback to calibrate the theory. The identification of common blends from a data-driven approach represents a first step to establishing emotional blends as concrete entities that summarize distinctive emotional experience. To complete the process of establishing emotional blends as meaningful entities, the blend extraction process needs to be accompanied by a demonstration of reliable inducibility for the emotional blends, to prove that data-driven blend extractions are not merely statistical artifacts. Without reliable experimental induction, the emotional blends cannot attain the same type of status as singular emotions have enjoyed in experimental and theoretical research.

As the topology of emotional blends gains clarity, the next major question in building an appraisal theory for emotional blends involves identifying the appraisals associated with the blends. The process of finding appraisals associated with emotional blends is modeled after singular emotion appraisal theory research, in which researchers ask participants to recall or relive a specific emotional experience while reporting appraisals associated with the experience (Smith & Ellsworth, 1985; Roseman, Spindel, & Jose, 1990). The recovery of appraisal profiles from both a data-driven emotional blend extraction and an experimental blend induction offers a systematic view to what are the appraisals supporting common emotional blends. Aside from the question over what are the cognitive foundations to the various blends, the appraisal profiles can also answer the question about how appraisals from different emotions become blended. With the appraisals profiles recovered summarizing a multitude of environmental implications within realistic situations, the appraisal profile of a blend can be examined against appraisals of each constituting emotion within a singular emotion setting, to show whether each emotion in the blend addresses the same environmental implications as it does on its own. If the role played by each emotion stays the same in a blend setting, an additive model of appraisal blending will be supported over a non-additive model. While apparently there are many more questions about the appraisals related to common emotional blends, finding out what are the appraisal profiles and how appraisals blend represents a first step forward to building an appraisal theory view to emotional blends.

The two questions push for the investigation into basic aspects of the emotional blend appraisal theory. I designed two studies to investigate these questions. The first study explored prevalent emotional blends with a data-driven approach. A set of exploratory clustering algorithms were used to extract several combinations of emotions from an existing dataset of emotional experiences. The cognitive foundations of the extracted emotional blends were examined through summarizing the appraisals related to each blend. The appraisal profile further led to the examination of whether appraisals from different emotions were blended additively or non-additively. The data-driven emotional blends and corresponding appraisals were compared with the theory-driven predictions shown in Table 2. The second study relied on emotional blend induction data to test the inducibility of a few blends, as well as to examine the organization of appraisals for emotional blends. In study 2, data were used in two separate analyses, each serving a specific purpose to answer the major questions. In the first part, the experimentally induced self-report emotions, along with the associated appraisals, were used to test if the targeted emotional blends were induced as intended. In the second part, the efficacy of emotional blends induction was further tested with voice samples gathered during induction. A machine learning algorithm was used to extract emotion information from the voice to offer an additional channel of evidence beyond self-reports.

CHAPTER 2

Study 1

Study 1 was an initial exploration into the world of emotional blends. With a pre-existing dataset that aggregated self-reported appraisals and emotions across many emotion elicitation scenarios, the aim for the first study was to extract commonly observed emotional blends based on self-reported emotions. From the extracted emotional blends, self-reported appraisal profiles related to each blend were summarized to describe the cognitive basis for that blend. The appraisal profiles were further analyzed to show how appraisals from each emotion were blended together. The projected results of study 1 consisted of the recovery of stable clusters of specific emotional blends obtained through two different clustering algorithms, with the appraisal profiles associated with the recovered blends to form an emotional blend summary table that resembles Table 2.

Method

Dataset

The dataset contained appraisal and emotion ratings across a wide range of emotion elicitation situations. Single studies typically relied on specific experimental situations or a pre-defined prompt to tap into a particular type or range of emotional experience. The limitation to do any meaningful exploration with data from single studies was the limited types of emotional experience that could be reasonably extracted: if the study never sampled a particular emotional blend experience, there was no chance that data exploration could find such emotional blend. Hence, to overcome the limitation of exploring data from a limited range of emotion elicitation situations, I intended to create a sample of emotion experience from many studies. In total, I aggregated data from 36 studies, to create a dataset that contains 3761 participants.

The emotional state reported by each participant was elicited with one of three methods: retrospection elicitation that asked participants to describe a specific emotional experience from the past, live studies that elicited emotions through specific experimental tasks such as mathematical problem-solving, and vignette-based studies that asked participants to imagine themselves in specific emotion-eliciting situations. The effectiveness of emotion elicitation across the different methods had not been very systematically studied, but one investigation by Robinson and Clore found comparability between retrospection and vignette based methods (2001). Depending on the study design, each participant might go through one or multiple rounds of emotion and appraisal assessments during the course of the experiment, such as baseline measures and post-induction measures. In studies with multiple rounds of assessments, I only retained appraisal and emotion assessments taken right after the emotion induction task, to sample the most elevated emotional experience. As a result, the aggregated dataset only contained one emotional experience for each participant, with no within-person dependency that might influence the independence of each appraisal and emotion assessment.

Participants responded to a series of appraisal and emotion questions on a 1 to 9 Likert scale. The specific emotions assessed varied somewhat across the 36 studies, but most studies assessed a common set of 9 emotions including anger, guilt, anxiety, sadness, hope, challenge, happy, pride, and gratitude. Across the studies, each of the 9 emotions had about 13.8% of missing data on average. This set of 5 positive and 4 negative emotions made up a large portion of the emotional experience people encounter, with each of the emotions offering a specific coping tendency that addressed a type of common environmental demand people could encounter. This set also formed the basis of the theoretical predictions for major emotional blends in Table 2. It should be noted that the 36 studies assessed more emotions than the set of 9 used for study 1. Those emotions not included in the study 1 analyses were not less important for defining emotional blends than those included. Rather they could be the defining elements in differentiating one blend from another. Nevertheless, the 9 individual emotions chosen here had received systematic reviews from emotion researchers (Smith & Lazarus, 1990; Smith, 1991), with adequate data availability across studies, hence proper for establishing a baseline to investigate emotional experience when the 9 emotions were blended.

Appraisals were also assessed across studies. The set of appraisals assessed in the studies were those proposed by Smith and Lazarus (1990). Unlike Table 2 which used congruence and incongruence appraisals to describe emotional blends, only a single congruence scale was used in the 36 studies. Although most of the appraisal variables had been covered in each study, there were a few studies that did not assess the full 7 appraisals. Across the studies, there were about 3.2% of missing data on average for each appraisal. These missing data did not pose serious challenges to identifying major emotional blends, since identifying the blends did not depend on the appraisal data. Instead, appraisals were associated with emotional blends in a descriptive fashion after the blends were identified. Nevertheless, to expand the availability of data, imputation was done through a k-nearest neighbor algorithm that searched for a complete data instance that best matched, for the variables they both shared, an instance with missing data. The missing data was then imputed with the values of the variables from the complete data instance. A comparison of descriptive statistics between the original variable and the imputed variable was shown for all appraisals and emotions in Table 3. Both the mean and standard deviation supported that the imputation procedure only changed the distributional form of the variables to a very small extent. To further explore the effect of imputation, I conducted a correlation of correlation analysis to examine the stability of relationship among appraisal and emotion variables prior to and after imputation. Specifically, I obtained all possible correlations among 7 appraisals and 9 emotions, essentially creating two 16 by 16 correlation matrices, one for the original data and one for the imputed data. Then a Fisher's r-to-z transformation was conducted to normalize both correlation matrices. The diagonal elements of 1s were removed before transforming the lower triangular half of each matrix into a 120 element vector. A Pearson correlation was then obtained between the two vectors, yielding a very high correlation of .998. This indicated that with imputation, how appraisal and emotion variables correlated among each other did not change much. As a result, I used the imputed dataset in study 1.

Table 3. Influence of data imputation on variable mean and standard deviation

	Set	Relevance	Congruence	Self	Other	Future	PFCP	AFCP	Anger	Guilt	Anxiety	Sadness	Hope	Challenge	Happy	Pride	Gratitude
N	Imputed	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761	3761
	Original	3757	3756	3610	3608	3335	3678	3753	3389	2921	2928	2925	3750	3752	3749	3151	2628
Mean	Imputed	6.84	4.32	5.75	4.84	5.49	5.03	6.61	4.02	3.19	4.97	3.71	4.79	4.77	3.63	3.28	3.95
	Original	6.84	4.32	5.74	4.85	5.53	5.00	6.61	4.19	3.37	4.87	3.96	4.79	4.77	3.63	3.37	4.17
Standard deviation	Imputed	2.08	2.58	2.50	2.64	2.14	2.38	2.16	2.80	2.36	2.53	2.63	2.62	2.52	2.90	2.59	2.74
	Original	2.08	2.58	2.55	2.69	2.24	2.39	2.17	2.85	2.56	2.71	2.79	2.62	2.52	2.90	2.70	2.99

Clustering Algorithm

The main objective of study 1 was to extract combinations of emotions that commonly occurred in emotion-eliciting situations. Two different clustering algorithms were used on the emotion ratings to identify the emotional blends that were common in the data set. Ideally, the clustering algorithms should yield clusters that were stable within each cluster and distinctive from other clusters. The first clustering method was Ward's method for hierarchical clustering, a method that combined samples of emotions into clusters based on their proximities to one another in an emotion rating space. The algorithm combined individual emotion samples into clusters, and clusters into larger clusters until only one cluster was left that contained all individual emotion samples. The decision of how many clusters to retain for the solution depended on indices that evaluated how compact each cluster was and how separate clusters were. Another criterion for deciding how many clusters to retain was the theoretical interpretability of new clusters: when a solution with N+1 clusters was compared to a solution with N clusters, high interpretability of the additional cluster would favor the solution with one more cluster. The second algorithm to be used was latent profile analysis (LPA). This model-based approach treated individual emotion samples as coming from different types of emotional experience with characteristic combinations of emotions, which were the emotional blends I was looking for in this study. Individual emotions sampled from the same emotional blend could have some variations around the blend average on each emotion. Model fit indices could help to decide how many groups, or emotional blends, should be retained to balance model fit and parsimony.

The two algorithms were chosen because they had some desirable characteristics that made them suitable for extracting emotional blends: Ward's method minimized changes of within cluster sum of square variance when combining clusters, while allowing for variation among instances within each cluster. Similarly, LPA grouped data points into specific classes with variability allowed within each class. For extracting prevalent emotional blends, I decided to allow for variation within clusters to capture the variability of emotional experience within each blend, as would be expected in reality. As various specific configurations could be used for each algorithm to yield different results, the results would likely be different between hierarchical clustering and LPA. Differences between algorithms were expected but not unwanted, as any valid clustering solutions should help with understanding the internal structure among emotions.

Both algorithms required some decisions to find a clustering solution that captured variability across clusters while not extracting too many clusters that ran the risk of over-fitting the data. For LPA, model fit would increase with more classes retained. However, more classes increased model complexity, reducing the generalizability of the model to datasets other than the one from which the model was fitted. Model fit indices such as AIC and BIC took into account the fit to data while penalizing for model complexity when comparing models with different numbers of classes. Model comparison tests like Lo-Mendell-Rubin Likelihood Ratio Test (LMR-LRT) could also help to determine the optimal number of classes through testing if a model yielded a significantly worse model fit than the model with one more class: if the fit was not significantly worse, the null hypothesis that the two models being compared offered similar fit would not be rejected (Lo, Mendell, & Rubin, 2001). Accordingly, the model with one less class was retained to favor model parsimony. For hierarchical clustering, the bottom of the dendrogram represented a solution that each data point was a cluster, whereas at the top all data points belonged to a single cluster. As a result, it was necessary to find a point to cut the dendrogram to retain a few clusters that meaningfully differed from each other. Internal measures of clustering solutions like Dunn's index and Silhouette index had been widely used to take into account the compactness of each cluster and the separation between clusters. Clustering solutions with high compactness within each cluster and high separation between clusters would have favorable high values on Dunn's index and Silhouette index.

Analysis Plan

Once LPA was conducted on the full sample, an optimal solution for the number of classes was obtained by evaluating the results of LMR-LRT from fitted models with increasing numbers of classes. Residual variance around the class means across different classes were allowed to differ, but not allowed to covary across different emotions within the same class. This restriction was applied because without it, the clustering algorithm would tend to under-extract classes with too much flexibility allowed. To avoid the risk of over-fitting models, the optimal number of classes obtained from the full sample LPA was conducted 10 times on a randomly drawn sample from the full dataset. Each randomly drawn sample contained 70% of the full dataset. Beyond visually assessing the extracted profiles from each randomly drawn sample, to assess the stability of profiles, I designed a stability statistic to quantify how likely the same two data points would be included into the same extracted profile across random sample draws. For example, random draw 1 and random draw 2 would have some overlapping data points, such as 5 for illustration. Of those 5 data points that were included in both draws, the lower diagonal matrix of a 5 by 5 table recorded if the two different data points were clustered into a shared profile across two draws. For illustrative purposes, I listed the 5 data points, denoted as D1 to D5, in the left panel of Figure 1, with the two rows reflecting the classification of D1 to D5 into different profiles across two draws. The classification results for each data point in one of the two draws were denoted with C1 to C4, with the subscript indicating which draw it belonged. The classification results were only relevant within each draw, and classes with the same number were not the same across draws. Data point 1 and 2 were classified into the same profile in draw 1, and were classified again into the same profile in draw 2, then the cell representing data point

1 and data point 2 in the right panel of Figure 1 would be noted by 1. If the two data points were classified into the same profile in draw 1 but not in draw 2, such as D1 and D3, the cell representing these two points would be noted by 0. In the case the two data points were classified into different profiles across both draws, even though the data points could be assigned into quite different profiles in each draw, such a situation still showed the classification was consistent with regard to these two data points. As a result, a 1 would be assigned to these two data points for measuring stability between draw 1 and draw 2. An example of this situation would be D3 and D5. In the end, the proportion of 1s in the lower diagonal quantified the stability of classification across two draws. A proportion was computed for each of 45 unique combinations of two sample draws. In the end, all 45 proportions were averaged to yield a single stability statistic that ranged between 0 and 1, with values closer to 1 indicating more stable classification solutions.

	D1	D2	D3	D4	D5
Draw 1	$C_1 1$	$C_1 1$	$C_1 1$	$C_1 3$	$C_1 4$
Draw 2	$C_2 1$	$C_2 1$	$C_2 2$	$C_2 2$	$C_2 1$

Draw 2 \ Draw 1	D1	D2	D3	D4	D5
D1					
D2	1				
D3	0	0			
D4	1	1	0		
D5	0	0	1	1	

Figure 1. Illustration of the stability index

For the Ward's hierarchical clustering analysis, first the full dataset was put through the clustering algorithm. Applied to a dissimilarity matrix, the Ward's method relied on minimizing within cluster variance when combining small clusters. It was one of the most widely used methods for cluster analysis, with variation allowed for data points within each cluster. To decide a cut point in the dendrogram produced with Ward's method, measures of both internal structure and stability were used to retain a small number of clusters. Candidate cut points were measured with Dunn's index and Silhouette index to quantify compactness within each cluster and separation between clusters. The cut point sporting the highest Dunn's index and the highest Silhouette index would be the best solution for the number of clusters to retain. The stability measure for candidate dendrogram cut points resembled what I proposed for the stability measures for LPA solutions: the proportion of overlapping cluster membership for data points across clustering solutions for different random samples drawn from the original dataset. The cut point that yielded the highest stability measure would present a preferred clustering solution. It was likely that the best solutions for different clustering algorithms would yield different numbers of clusters, and potentially clusters with qualitatively different interpretations. Subjective judgments would be required to decide which clustering solutions to retain, especially if there was not a solution that was clearly the best in terms of model fit and stability. Following the selection of a clustering solution, appraisals associated with each cluster would be summarized by taking the average for each appraisal variable from all data instances categorized into the cluster. The average appraisals associated with each blend were further examined for whether their levels were commensurate with the appraisals traditionally accepted for each of the prominent emotions within that blend.

Hypotheses and Predictions

Study 1 has no specific hypothesis to be tested due to its exploratory nature. The potentially massive range of emotional blends that could be observed made it very difficult to produce specific predictions for how many prevalent blends would be recovered or what they would look like. However, existing theoretical research does offer clues as to likely blends based on how singular emotions are organized relative to each other (Robinson & Clore, 2001). Although the blends proposed from Table 2 are only based on the limited information from previous research, study 1 represents a data-driven approach that could serve as a validation of the blends obtained through theoretical review.

Results

In this section, the results for study 1 are organized in to three major sections. First, a detailed account of the model selection process examined various candidate clustering solutions from the two clustering methods. The process involved selection based on both different model fit indices and model interpretability. Next, once the final clustering model had been selected, the empirically recovered blends were compared to the theoretically predicted blends of Table 2. Finally, the associated appraisals for each emotional blend were computed, and compared to the appraisals related to the emotions with the emotional blends.

Model Selection

For each classification method, I fitted solutions between 2 to 25 classes to find an optimal solution. The reason for going up to 25 classes was not to explore the potential fine-grained differentiation of emotional blends at such a high number of classes: trying to consider so many clusters simultaneously would be extremely difficult and likely noisy. Instead, the high number of classes provided a more exhaustive search space allowing for greater confidence in the solution ultimately deemed optimal. An analogy to exploring a wider number of classes would be to use different starting values in iterative maximum likelihood estimations to avoid local maxima.

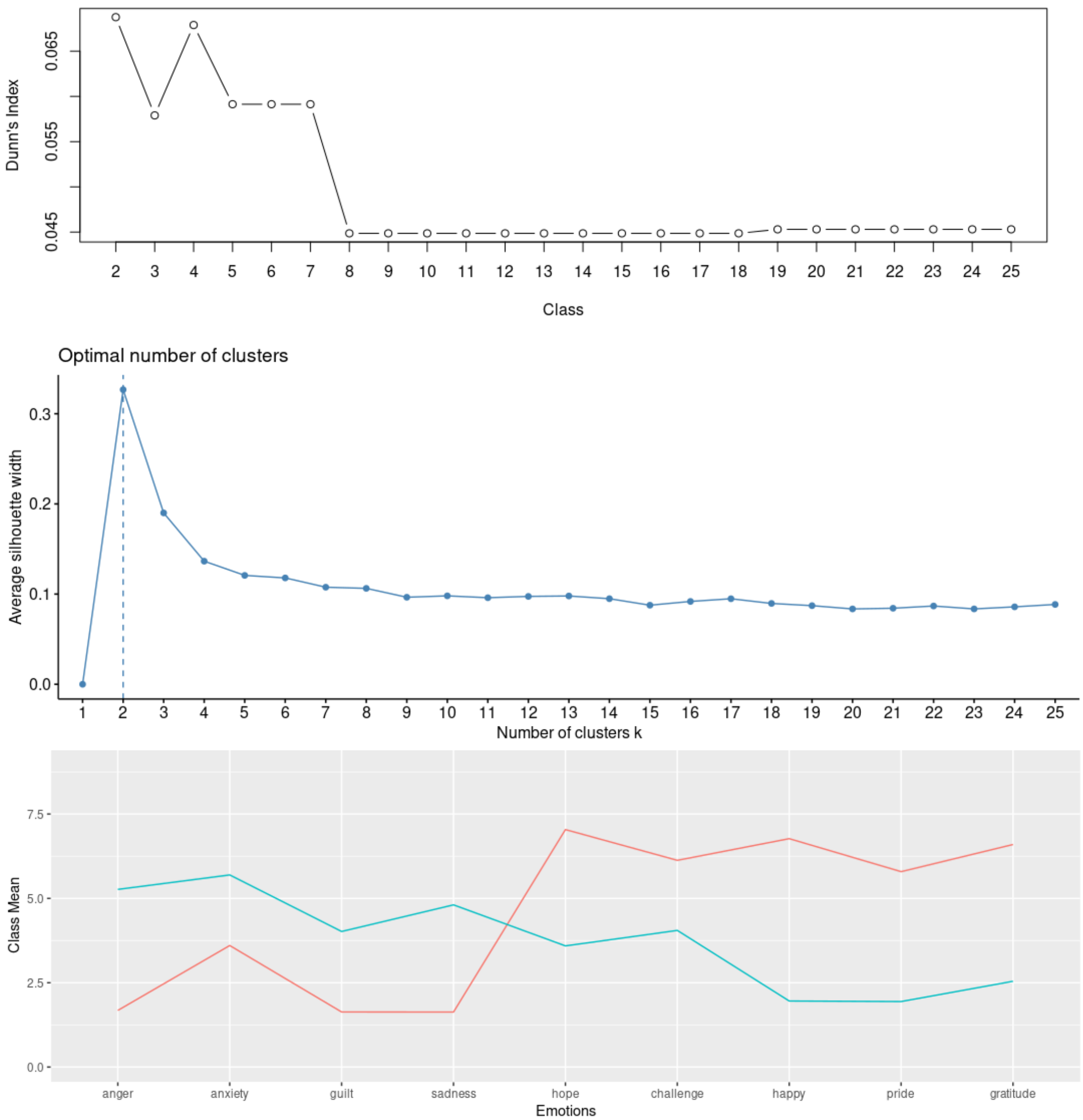


Figure 2. Dunn's index, Silhouette index, and emotion profiles of 2-class solution

For the Ward's method, the dendrogram computed from Euclidean distance was cut at different points to yield solutions between 2 to 25 classes. Dunn's index and Silhouette index were computed for each solution. As can be seen from the top panel of Figure 2, these indices were not useful for identifying the optimal number of classes: both favored a 2-class solution, which was essentially recovering a positive and a negative cluster of emotional experiences, as seen in the bottom panel of Figure 2. These two indices, because of their reliance on the distance measures to determine if the within class distance is smaller than the cross-cluster distance, did not function well in high dimensional space where distance among different data points became similar disregarding which dimensions contributed to the distance (Tibshirani, Walther, & Hastie, 2001). In order to deal with this situation, which was unexpected at the analysis planning stage, I decided to ignore the two fit statistics but relied on the stability index to find a proper solution for the Ward's method, and then inspect the recovered classes for interpretability. While the various fit indices are commonly accepted in research involving clustering, such indices can be problematic in specific situations where more flexible judgment is required. On the other hand, stability, interpretability, and the size of cluster all contributed to the quality of solutions that were no less meaningful to my research topic than the compactness of clusters that were quantified by the established fit indices.

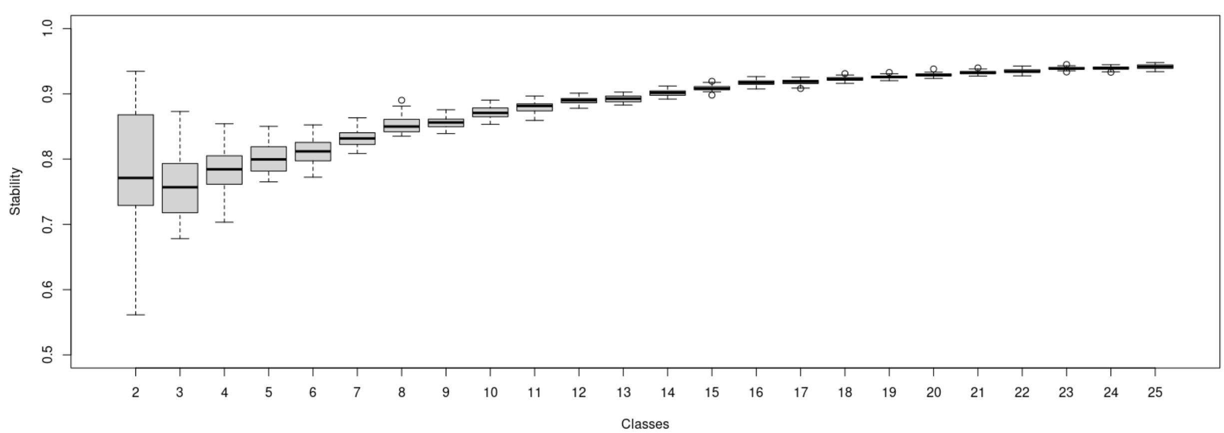


Figure 3. Ward's method stability index

Figure 3 illustrates the stability of the Ward's method solutions for the different number of classes. The median stability index, as well as the interquartile range for the stability index, computed across all possible 45 pairs from the 10 resampling iterations are shown for each number of classes in the form of a box-and-whisker plot. From Figure 3, an upward trending stability index is obvious with increasing numbers of classes. This presented an impasse: clearly the classification became more stable with more classes, hence an optimal solution based on the stability index would reach an extremely high number of classes. This observation pointed to the lack of prominent clustering signal with the Ward's method for hierarchical clustering, otherwise there should be solutions with stability that breaks from the trend shown in Figure 3. While the nature of this thesis did not warrant a full investigation into the mathematical properties of the stability index and its relation to the number of classes, I speculated strong clustering information should be picked up by the stability index through the resampling approach. The apparent lack of clustering information could possibly be due to an actual lack of significant significant clustering information, which could be corroborated with the other clustering method, or it could reflect that the Ward's method was not fully appropriate for these data due to their high dimensionality. The next step was to look at the LPA clustering method to examine the two possibilities.

For LPA, both LMR-LRT and the stability index were used to explore solutions ranging from 2 to 25 classes. I chose the mclust package in R to fit the LPA model, with the "VII" covariance structure chosen to allow for unequal variance across classes but not covariances (Scrucca, Fop, Murphy, & Raftery, 2016). LMR-LRT was obtained with the TidyLPA package in R (Rosenberg et al., 2018). The full dataset was used for using LMR-LRT to select the best number of classes. For LMR-LRT, the first time a nonsignificant result appeared when a 19-class solution was compared against an 18-class solution, hence presenting an optimal solution of 18-classes. Because an 18-class model would be very difficult to interpret, it was not pursued further. The stability index, presented in Figure 4, decreased with more classes before 6 classes, after which the stability index trended upwards. Notably, the 10-class solution broke away from the increasing trend between 6 and 25 classes by exhibiting an especially high stability index. I interpreted this solution as capturing a prominent clustering signal where other solutions had their stabilities conforming to a general trend. In other words, the sudden upshift in stability at 10 classes could be viewed as a combination of increasing stability as a function of the number of classes, and a particularly strong classification signal at 10 classes.

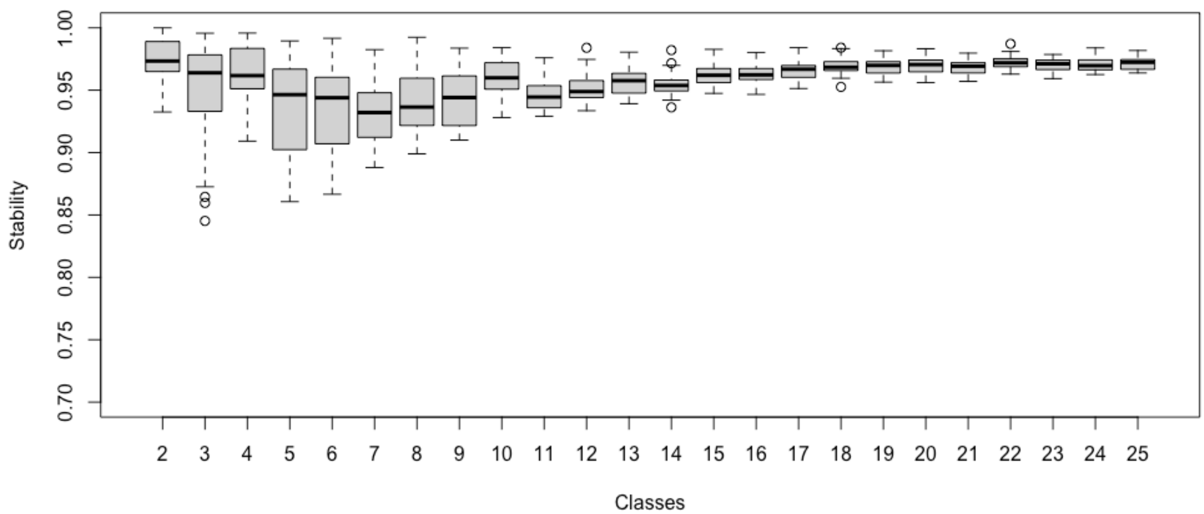


Figure 4. LPA stability index

Parallel to the statistical indices, the interpretability of the clustering signal within data presented another crucial criterion to the usefulness of the clustering solutions. Indeed, if clusters found with a certain method were not meaningfully interpretable, the clusters would likely represent methodological noise. Hence, I decided to examine the correspondence between LPA and Ward's method, at different numbers of classes. A high correspondence across methods would demonstrate that the recovered clustering structure was prominent enough to be picked up by methods that approached clustering differently. I examined the correspondence between 2 to 12 classes. Solutions above 12 classes were excluded due to low interpretability. Chord diagrams offered a visual aid to understand how classes from one method matched with those from the other method. This tool was particularly useful in that it allowed for tracking classes from one method breaking and merging into classes from another method both in terms of direction and quantity. For each number of classes, a chord diagram was used to examine the correspondence between the two clustering methods. Figure 5 portrays the chord diagram created with 10-class solutions, one of the 11 diagrams I used for correspondence checking. The full set of 11 diagrams are presented in Supplementary Figure 1.

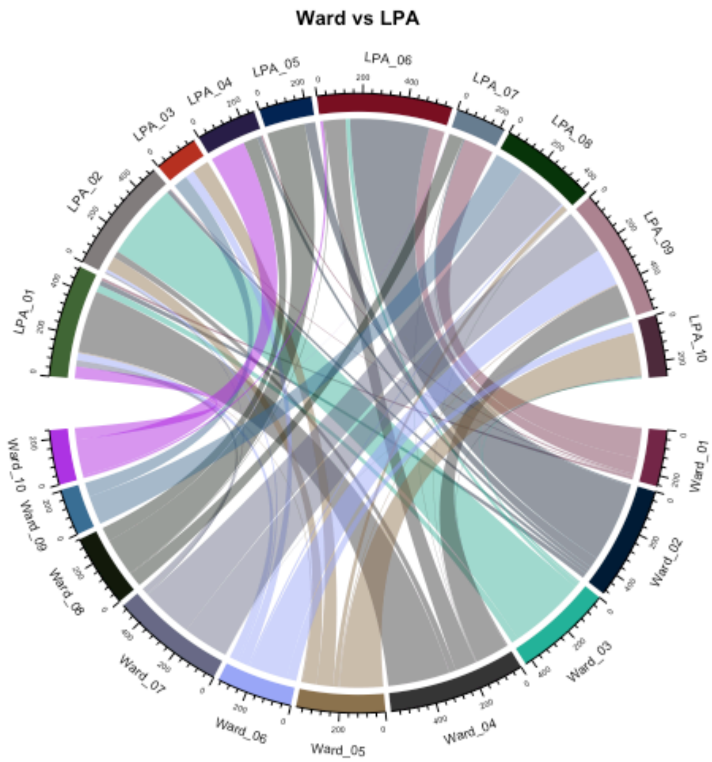


Figure 5. Chord diagram for 10-class solution

For the 2-class solution the class correspondence between the two methods was almost perfect. However, the two classes merely represented a blend of all positive emotions and a blend of all negative emotions, respectively, and thus were not interesting. The correspondence for 3-classes solutions reduced considerably with classes from the Ward's method solution splitting into multiple LPA classes. Similarly, LPA classes could be traced to multiple Ward's method classes without a predominant source. From 4 classes onward, although correspondence appeared to improve gradually, clear assessment of correspondence became challenging with more complex diagrams. To offer a straightforward comparison of correspondence, the stability index used for determining the optimal number of classes for LPA was adapted to capture the stability between the LPA solution and the Ward's method solution of the same number of classes. Because LPA relied on random seeds for determining model estimation starting values, the classification solutions changed slightly each time the model was fitted on the same dataset. To better capture a representative correspondence for a certain number of classes, the LPA model was fitted 10 times on the full dataset with random starting values, whereas the Ward's method was fitted once. The stability index was computed between each LPA model fitting and the Ward's method solution. Figure 6 portrays the correspondence between the two methods from 2 class to 20 classes. For each number of classes, each boxplot in Figure 6 summarizes the 10 stability indices computed between each of the 10 LPA model fittings and the Ward's method solution. As seen in Figure 6, correspondence increased from 3 classes, with the largest jumps occurring at 6-classes, 9-classes, and 10-classes. The slow and gradual increase in correspondence post 10-classes was not interesting as such increase was likely a mathematical property of the stability index, nor were those solutions with very high number of classes meaningfully interpretable. As a result, the 10-class solutions presented the last prominent signal for correspondence increases, which combined with the prominent stability signal for the 10-class LPA solutions, led me to retain the 10-class LPA solution as the final model for the predominant emotional blends within the study 1 dataset. This model is examined in detail in the analyses that follow.

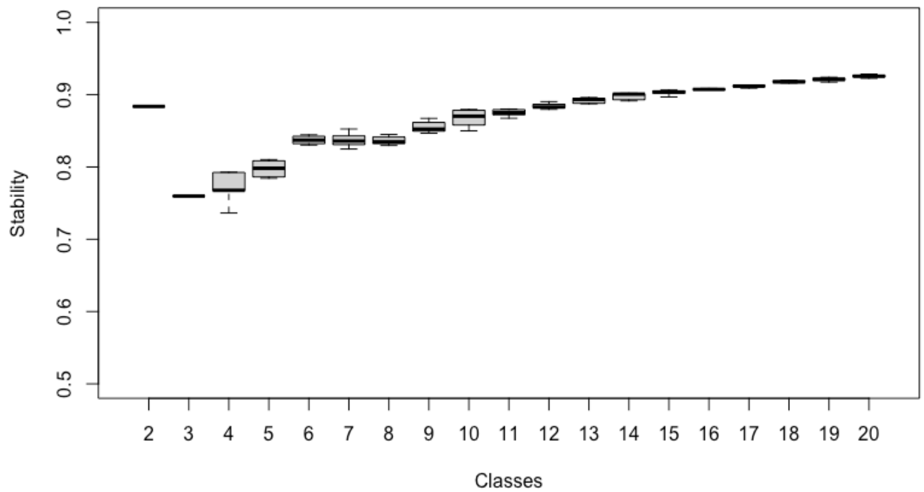


Figure 6. Correspondence between LPA and Ward's method solutions

Extracted Emotional Blends

Figure 7 presents the class averages of the nine emotions for each of the 10 recovered classes. These profiles are presented in 3 panels for clarity. The first class had medium-to-high levels of anxiety, hope, and challenge, and was comprised of 494 emotional episodes. Other emotions in this class all hovered around the medium level. This class best resembled the opportunity blend predicted in Table 2, but at a medium level of intensity. Together, the emotions in class 1 represented people in an anxiously expectant state. In contrast to class 1, class 2, with 506 episodes, had low ratings on virtually all emotions, with the only exception of a slightly elevated anxiety rating. Class 2, which could be described as a relatively non-responsive blend, did not resemble any of the predicted emotional blends. Class 3 was characterized by each of the negative emotions, except guilt, being of high intensity. The positive emotions were all low intensity. This blend closely resembled the anger blend projected in Table 2 which included high anger, sadness, and anxiety. This class accounted for 192 emotional episodes.

Figure 7. Emotion profiles for 10 LPA classes

Class 4 had 265 episodes, with high positive emotions throughout. However, this blend also had prominently elevated anxiety, which differentiated class 4 from class 5, where all positive emotions were high but with hardly any elevation among the negative emotions. Class 5 accounted for 232 emotional episodes. Among the smaller differences, class 4 had higher challenge than class 5, which echoed the difference between the two classes on anxiety.

Furthermore, class 5 had a slightly higher happiness rating than class 4. Class 4 appeared to represent an anxious excitement emotional blend that might keep people on the lookout for potential obstacles even within a positive situation, where class 5 appeared to describe a purer form of excitement. Class 6 presented another seemingly expectant state, somewhat milder than the one presented in class 1. This blend accounted for the most emotional episodes in the dataset, with 593 data points attributed to this class. This class had medium to high positive emotions with elevated hope and challenge. It differed from class 1 in its depressed negative emotions and much less prominent anxiety. This blend closely resembled the mild positivity blend as predicted in Table 2.

Class 7 with 221 emotional episodes had high hope, happiness, and gratitude. Challenge and pride were depressed compared to other positive emotions. This class had low negative emotions. In this class, challenge and hope, which had matched each other in intensity such as in class 1 and 6, had a large discrepancy between them. This class closely resembled the predicted other-benefit blend with high gratitude and happiness. Class 8 had high negative emotions across the board, and very low positive emotions overall. This class had the highest guilt level among all the 10 classes. Class 8, which accounted for 423 emotional episodes, best resembled the self-negativity blend predicted in Table 2. Class 9, with 565 episodes, had high anger, anxiety, sadness, and challenge. Guilt and hope were also somewhat elevated. The guilt level in this class was in between that in class 3 the anger blend and class 8 the self-negativity blend. This middling level of guilt, combined with the high anger, meant that class 9 best resembled the predicted dual negativity blend, which combined the source of harm from the anger blend and the self-negativity blend. Predicted to be elevated in the dual negativity blend, the moderate levels of challenge and hope seen in class 9 indicated that there were environmental obstacles that needed and could be addressed. Different from all other classes, class 10 had the look of a singular emotion state, rather than a blend. This class, accounting for the remaining 270 emotional episodes, had only anger as a prominent emotion. Class 10 the anger-only class and class 3 the anger blend both featured high anger, but class 3 had other elevated negative emotions accompanying anger.

Associated Appraisals

The appraisals associated with each class of emotional blend took the class average on each of the 7 measured appraisals. First, a MANOVA helped to establish a statistically significant difference for all appraisals across the LPA classes, $F(63, 26257) = 67.763$, $p < .001$; Pillai's trace = 0.979. A series of ANOVAs was then conducted for each of the appraisals. As documented in Table 4, there were significant differences across classes for each appraisal.

Table 4. ANOVA results for appraisals across LPA classes

Appraisal	df	F	p
Relevance	(9, 3751)	90.3	<.001
Congruence	"	539.2	<.001
Self-accountability	"	58.38	<.001
Other-accountability	"	27.88	<.001
Future expectancy	"	87.97	<.001
PFCP	"	63.97	<.001
AFCP	"	86.19	<.001

Figure 8 depicts the average appraisals for each class. The 10 classes are broken down into 3 panels for clarity. In describing the appraisal profiles, I drew on the singular emotion appraisal-emotion relation listed in Table 1 to evaluate how the appraisals profiles of the LPA classes supported the emotion profiles. Class 1 the opportunity blend had moderate levels of anxiety, hope, and challenge. The appraisals associated with each of the constituting emotions included high relevance and low congruence for anxiety, hope, and challenge, high future expectancy for hope and challenge, high PFCP for challenge, and low AFCP for anxiety. The appraisals for the opportunity blend included high relevance and medium congruence, as well as moderate levels of future expectancy, PFCP, and AFCP. The AFCP level, which had an elevated rating across all classes, occupied a medium level compared to other classes. High relevance in the blend matched the high relevance typically associated with anxiety, hope, and challenge. While typically these three emotions have low congruence, the medium congruence level in the blend meant that they would not have high intensity. The moderate future expectancy corresponded to what was typically associated with hope and challenge, in weak forms. The moderate PFCP and AFCP in the blend aligned with the key coping potential appraisal for challenge and anxiety, respectively, again in weak forms. In class 2 the non-responsive blend, the relevance and congruence appraisals were lower than those of class 1. The relevance appraisal for this class was the lowest among all classes. This blend also had high AFCP. Anxiety, the only emotion that had a slight elevation in this blend, was typically associated with high relevance, low congruence, and low AFCP. The low relevance, a moderate level of congruence, and high AFCP in this blend likely limited the intensity of anxiety. Class 3 the anger blend had high anger, anxiety, and sadness. The appraisals associated with each of the constituting emotions included high relevance and low congruence for all the emotions, high other-accountability for anger, low PFCP for sadness, and low AFCP for anxiety. The appraisal profile for this blend included high relevance, low congruence, elevated other-accountability, low PFCP, and low AFCP. The relevance and congruence appraisals in the blend aligned with the relevance and congruence levels theoretically expected for the three negative emotions this blend. The high other-accountability expected for anger was consistent with the high other-accountability in the blend. The low PFCP in the blend aligned with what was expected for intense sadness. The low AFCP in the blend was consistent with what was expected for intense anxiety. In the first three classes, the appraisals for each of the constituting emotions in the blends were retained in the overall blend appraisal profiles, thus lending support to the additive model of appraisal organization in blends.

Figure 8. Appraisal profiles for 10 LPA classes

Class 4 the anxious excitement blend had high hope, challenge, happiness, pride, gratitude, and elevated anxiety. The appraisals associated with each of the constituting emotions included high relevance for all the emotions, high congruence for happiness, pride, and gratitude, low congruence for hope, challenge, and anxiety, high self-accountability for pride, high other-accountability for gratitude, high future expectancy for hope and challenge, high PFCP for challenge, and low AFCP for anxiety. The appraisal profile for class 4 consisted of high congruence and relevance, the combination of which aligned with what was expected for high happiness, pride, and gratitude. Class 4 also had high self-accountability relative to other-accountability, although this accountability setup did not lead to high pride over gratitude. Class 4 had high future expectancy and PFCP compared to other classes, but it did not have relatively high AFCP. The high future expectancy and PFCP corresponded to what would be expected for high hope and challenge, while the lower AFCP corresponded to an elevated anxiety rating. Although both hope and challenge were typically associated with low congruence, they both had high ratings despite the high congruence in class 4, a patterned also observed in class 5 and class 7. Class 5 the pure excitement blend had high ratings on all the prominent emotions from class 4 except for an elevated anxiety. The two blends shared many appraisals including a bias toward self-accountability that somehow mapped onto high gratitude. The most perceivable difference between class 4 and 5 in appraisals was the higher AFCP in class 5, a distinction that was reflected in the lower anxiety in class 5. Class 6 the mild positivity blend had elevated but not intense overall positive emotions and a slightly elevated anxiety, similar to class 4 but with all emotion intensity greatly reduced. For the appraisals, class 6 had medium levels of relevance and congruence, both of which limited the intensity of the positive emotions as well as anxiety.

Besides low relevance, the higher AFCP in class 6 compared to class 4 possibly contributed to the lower level of anxiety in class 6 than class 4. Following the additive model, the appraisals for the constituting emotions in class 4 to 6 were largely retained in the blend appraisal profiles, but the high anxiety associated with a high congruence in class 4 marked a departure from| the additive model. Also contradicting the additive model were hope and challenge in class 4 and 5 that deviated from their traditionally accepted low congruence appraisal, as well as high gratitude in class 4 and 5 that was associated with high self-accountability instead of high other-accountability.

Class 7 the gratitude blend had high hope, happiness, and gratitude. The appraisals associated with each of the constituting emotions included high relevance for all the emotions, high congruence for happiness and gratitude, low congruence and future expectancy for hope, and high other-accountability for gratitude. The appraisals of this blend included high levels of relevance, congruence, and other-accountability. Both happy and gratitude were consistent with the observed high relevance and congruence in the blend, whereas the high other-accountability was consistent with what was expected for gratitude. This class also had higher future expectancy and AFCP relative to PFCP. The high future expectancy was consistent with elevated hope whereas the depressed PFCP was consistent with low challenge in the gratitude blend. Class 8 the self-negativity blend had high anger, guilt, anxiety, and sadness. All four negative emotions were theoretically associated with high relevance and low congruence. Because of the high guilt, the high anger in this blend was likely the self-directed anger (Ellsworth & Tong, 2006). Both guilt and the self-directed anger had been associated with high-self accountability. Other appraisals relevant to the constituting emotions included low PFCP for sadness and low

AFCP for anxiety. Class 8 had high relevance, low congruence, high self-accountability, low future expectancy, low PFCP, and low AFCP. The high relevance and low congruence were aligned with the appraisals hypothesized for the intense negative emotions. The high self-accountability was consistent with what was expected for the high guilt and self-directed anger. The low PFCP observed in the blend was consistent with the expected PFCP for the high sadness.

Compared to class 8, class 9 the dual negativity blend had lower guilt, anxiety and sadness, as well as higher hope and challenge. Compared to class 8, the appraisal profile for class 9 had higher congruence, future expectancy, PFCP, and AFCP. Class 9 also had a smaller distinction between self and other-accountability than class 8. The higher congruence was consistent with theoretically hypothesized appraisals for the lower anxiety and sadness, while the less bias toward self-accountability in the blend aligned with the lower self-accountability associated with a lower guilt. The increased future expectancy corresponded to what was typically associated with higher hope and challenge, while the increased PFCP aligned with what was expected for the higher challenge and lower sadness. Class 10, the anger-only class, had only medium level of relevance, diverging from the high relevance appraisal typically associated with anger. This class also had low congruence, as well as high other-accountability in comparison to self-accountability, both consistent with theoretically hypothesized congruence and accountability appraisals for anger. Compared to the anger blend class 3, class 10 had pronouncedly higher future expectancy, PFCP, and AFCP, which limited the intensity of anxiety and sadness. For the final 4 blends, the additive model of appraisal organization in blends was supported by most of the emotions except for the high congruence associated with hope in class 7, as well as the medium level of congruence associated with high anger in class 10.

Discussion

The results from study 1 demonstrated the existence of distinctive emotional blends that were consistently observed across a broad sample of emotional experiences. The 10-class solution from LPA was able to demonstrate high stability in the clustering structure relative to the number of classes retained. Besides the anger-only class, each of the emotional blend extracted from LPA described a unique emotional state consisting of elevations in multiple emotions. A majority of the recovered emotional blends could be matched with the theoretically predicted emotional blends. The average appraisals associated with each extracted emotional blend formed distinctive appraisal profiles fundamental to the elevated emotions in the blends. Appraisals for the constituting emotions of most emotional blends followed the traditionally established singular emotion appraisal-emotion relations, although there were blends where appraisal-emotion relations changed based on what other appraisals and emotions were present in the blends. I would like to make a few more elaborations.

Study 1 recovered many blends predicted in Table 2. In Table 5, the blends predicted prior to this study are reproduced, each presented along with the LPA blend that best matches the predicted blend. Among the 9 blends predicted, 6 have close matches from the 10-class LPA solution. Most of the predicted blends with predominantly negative emotions were recovered from the LPA solution, with the exception of the anxiety blend. For the LPA classes, there does not exist any class that features high anxiety and sadness without anger, as was predicted for the anxiety blend. Class 1 with an anxiously expectant emotional state, is the closest to the anxiety blend, although a lack of match on sadness moves class 1 closer to the predicted opportunity blend. In the extracted blends with high anxiety and sadness, such as in the anger blend, the self-negativity blend, and the dual negativity blend, there are always emotions like anger or guilt indicative of sources of accountability. It appears that a lack of an anxiety blend may be explained with that loss and motivation to address a threat appear to cooccur when there are clear sources of wrongdoing, either from self-blame or other-blame. Among the predicted blends that largely feature positive emotions, a self-benefit blend is absent from the LPA solution. The extracted blends most similar to the self-benefit blend are class 4 the anxious excitement blend and class 5 the pure excitement blend, although both blends feature other positive emotions like hope, challenge, and gratitude. While class 4 and 5 do not have clear mappings onto the predicted blends, the coupling of pride and gratitude in these classes perhaps points to a shift in what gratitude signals, likely a high gratitude not directed toward other people but toward impersonal factors like luck (Teigen, 1997). This interpretation has support from the appraisal profiles associated with these two blends, which do not have high other-accountability. If the observation of high gratitude is explained this way, both class 4 and 5 can be matched with the predicted self-benefit blend: both classes had high relevance, high congruence, and high self-accountability, as predicted for a self-benefit blend.

Table 5. Predicted emotional blends and corresponding LPA classes

Blend	Emotions in blend	Appraisals	Closest LPA class	LPA class mean appraisals
Self-negativity blend	Guilt, anger (self-directed), sadness	High relevance Low congruence High incongruence High self-accountability Low future expectancy Low problem-focused coping	Class 8	High relevance Low congruence High self-accountability Low future expectancy Low problem-focused coping Low accommodative-focused coping
Dual negativity blend	Guilt, anger (self-directed), anger (other-directed), sadness, anxiety, hope	High relevance Low congruence High incongruence High self-accountability High other-accountability Medium future expectancy Low problem-focused coping	Class 9	High relevance Low congruence Medium to high self-accountability Medium other-accountability Medium future expectancy Medium problem-focused coping
Anxiety blend	Anxiety, sadness, challenge	High relevance Low congruence High incongruence Medium future expectancy Medium problem-focused coping Low accommodative-focused coping		
Anger blend	Anger (other-directed), sadness, anxiety	High relevance Low congruence High incongruence High other-accountability Low future expectancy Low problem-focused coping Low accommodative-focused coping	Class 3	High relevance Low congruence Low self-accountability High other-accountability Low future expectancy Low problem-focused coping Low accommodative-focused coping
Mild positivity blend	Calm/Tranquility, mild other positive emotions	Low relevance Medium-to-high congruence Low incongruence High future expectancy.	Class 6	Medium relevance Medium congruence Medium to high future expectancy High accommodative-focused coping
Opportunity blend	Challenge, hope, anxiety	High relevance Low congruence High incongruence High future expectancy High problem-focused coping Low accommodative-focused coping	Class 1	High relevance Medium congruence Medium to high self-accountability Medium to high future expectancy Medium problem-focused coping Medium accommodative-focused coping
Self-benefit blend	Happiness, pride	High relevance High congruence Low incongruence High self-accountability		
Other-benefit blend	Happiness, gratitude	High relevance High congruence Low incongruence High other-accountability	Class 7	High relevance High congruence High other-accountability High future expectancy Medium problem-focused coping High accommodative-focused coping
Bittersweet blend	Happiness, sadness	Medium relevance High congruence High incongruence Low future expectancy Low problem-focused coping		

In contrast to the single mild positivity blend predicted in Table 2, a notable finding from study 1 is the existence of multiple mild emotional blends including class 1 the opportunity blend, class 2 the non-responsive blend, and class 6 the mild positivity blend. The mild emotional blends do not include intense emotions of any kind, a set up that may prepare people to attend to potential challenges while not exhausting them. The three extracted mild emotional blends all have their distinctive appraisal and emotion profiles. Class 1 the opportunity blend shows a considerable portion of participants occupying an emotional state with elevated anxiety, hope, and challenge relative to other emotions, albeit only at a medium level. The appraisal profile of this blend has medium level of congruence, as well as elevated but not high levels of self-accountability and future expectancy, all of which are less extreme than those previously predicted for an opportunity blend. As a result, the opportunity blend recovered from data conveys a somewhat changed emotional blend from the the one predicted, in a far less dramatic picture of people cautiously proceeding in normal situations in anticipation of potential challenges. Compared to the other two mild emotional blends, class 1 has higher relevance appraisal and lower AFCP, both of which may explain its higher anxiety than the other two blends. The high relevance also means that the opportunity blend appears to prepare people to engage potential challenges once they are already in a clear goal-oriented state. As people navigate such situations, elevated anxiety, hope, and challenge may help people to avoid potential danger, while motivating them to persevere and actively engage effort in their pursuit of goals (Smith, 1991).

Class 2 the non-responsive blend has the lowest relevance appraisal among all the classes from LPA, which corresponds to a low emotion intensity across the whole spectrum of emotions relative to class 1 the opportunity blend. While it can be argued that participant response tendencies led to this emotional blend, the appraisal rating responses from the same group of participants are not uniformly low, especially for AFCP, hence indicating that the low emotional ratings capture a true lack of emotions in this blend. The high AFCP limited the intensity of anxiety as compared to class 1. The slight elevation of anxiety in the non-responsive blend requires further discussion. In positive emotional blends like class 5 the pure excitement blend and 7 the other-benefit blend, as well as in class 6 the mild positivity blend, there is always a slight elevation in anxiety. A slight anxiety, even in the most positive situations, may the serve the role of quickly engaging with any potential danger, a mechanism perhaps rooted within the evolutionary significance of emotions for detecting danger and promoting survival with adaptive behaviors (Smith & Lazarus, 1990). From this perspective, a common mild anxiety observed widely within the non-responsive blend and the positive emotional blends seen in this study, should have helped to blur the lines between the mild emotional blends and blends that prominently feature specific emotions.

Class 6 the mild positivity blend differs from the focused state of the opportunity blend by the absence of a strong relevance appraisal. Without a clear goal to work toward, as well as a high perceived capability for accepting any potential turn of events, people in the mild positivity blend may not need high anxiety to help them navigate obstacles to goal seeking, because there is no goal to seek in the beginning. The higher congruence appraisal in the mild positivity blend relative to the opportunity blend plays a role in depressing the negative emotions overall. Additionally, hope and challenge in the mild positivity blend appears to differ from their role in helping people sustain in face of obstacles, as described by Smith (1991). Hope and challenge in the mild positivity blend are not influenced by general positivity like in class 4 or 5, since happiness rating is lower than hope and challenge in class 6. One possible interpretation is that with the absence of threat, these two emotions arise to prepare people for taking on potential opportunities and looking toward personal growth, along the line of the broaden-and-build theory for positive emotions (Fredrickson, 2004). In any case, the observation of multiple mild emotional blends paints a nuanced picture of those emotional states that typically go unnoticed, but are nevertheless prevalent in the emotional life of people.

Class 10 the anger-only class requires some further discussion. This class, with only high anger, supplements the emotional blend view: although emotional blends are largely omnipresent in the emotional life of people, not all emotional experiences involve multiple emotions. The anger-only class supports that in certain circumstances, strong and singular emotions can appear by themselves to capture the whole adaptive need for the emotions to address. Compared to the anger blend, the anger-only class has only a medium level of relevance appraisal and high AFCP, both of which suppress strong anxiety. However, this medium level of relevance, combined with high other-accountability is still able to elicit very high anger. Besides the environmental implication that leads to pure anger, there are no other apparent situational implications: people in this class appear to not feel a need to act to resolve challenges or mourn for losses, nor are there perceived threats. With the sole focus on the accountability of others, and a less on personal relevance, people may be feeling some type of righteous indignation about injustice although there are perhaps no direct and immediate implications for them.

The appraisals profiles associated with the extracted emotional blends were able to explain most of the elevated emotions within those blends, indicating that when emotions are blended, the associated appraisals characteristic to each singular emotion can be retained within most circumstances. This overall observation across the blends offers strong support to the additive view of how appraisals function within emotional blends. However, support to an additive view does not mean that all types of emotions can be freely combined regardless of context. The lack of an anxiety blend that has only anxiety and sadness from the LPA solution is one example: as shown in the anger blend or the self-negativity blend, appraisals associated with anxiety and sadness in their singular forms can be additively combined when there are other emotions with clear sources of accountability, but that does not mean anxiety and sadness can be combined by themselves according to the LPA blends. The extracted emotional blends and the appraisal profiles associated with the blends offer a look at the types of complex situational implications that give life to emotional blends, but contextual constraints that allow or block the formations of specific situational implications require future investigation to better understand the applicability of the additive model of blending.

Simultaneous to the broad support for the additive view are the multiple appraisal profiles of emotional blends not conforming to the appraisals typically associated with some of the emotions constituting the blends. One notable divergence occurred in class 4 the anxious excitement blend, where a high congruence appraisal was associated with high anxiety amid other intense positive emotions. In this blend, anxiety in the context of intense positive emotions no longer functions as an indicator to threat that corresponds to low perceived congruence, as it does in its singular emotion form. Compared to class 5 the pure excitement blend, class 4 has a lower AFCP level, which may be the source of the elevated anxiety in this blend. Despite a high congruence, which typically reduces anxiety, a reduced AFCP exerts a powerful influence on anxiety. The example of anxiety suggests that in certain situations, key appraisals typically associated with certain emotions may be no longer necessary. A key question emerges: is anxiety in class 4 still the anxiety that responds to a personally relevant threat? When combined with intense positive emotions, anxiety could not have been serving as a threat-responding mechanism if there is no threat present. Instead, subtleties in situational implications may help to determine the specific nature of anxiety in this case. In class 4, the higher other-accountability, lower future expectancy, and lower PFCP compared to class 5 the pure excitement blend may all have indirectly contributed to the elevated anxiety. Although these three appraisals have not been typically linked to anxiety, in a positive situation where others have more involvement, they may indicate that, although things are going well, the person is concerned that this could change. Thus the specific combination of the three appraisals, combined with the high congruence and lowered AFCP, appears to have created a hypothetical threat that led to elevated anxiety.

While the contextual influence on the variability in anxiety is a speculative interpretation at this stage, the functional variability of emotions due to changes to key appraisals is certain. Previous studies have documented variabilities of specific emotions such as the self-directed anger that lack a high other-accountability typically associated with anger (Ellsworth & Tong, 2006). The emotional blends extracted in this study help to systematically expand the list of emotion variants by first identifying the commons blends and then finding occasions where key appraisals are not observed for the blended emotions. Challenge and hope have traditionally been associated with low congruence, as is the case for the dual negativity blend, where the two emotions serve sustainer roles to motivate people addressing present obstacles. However, in other situations such as in the mild positivity blend, hope and challenge may serve preparatory roles for potential opportunities, when there is no low congruence due to obstacles. Gratitude, typically associated with high other-accountability, occurs in blends with low other-accountability such as in the anxious excitement blend and the pure excitement blend. In these blends, gratitude is not directed toward others due to low other-accountability, unlike how it functions within class 7 the gratitude blend. Anxiety, challenge, hope, pride, along with anger which has been studied previously for its functional variabilities, can all be elicited to serve situation-specific functions when some key appraisals are not present, at least in the context of particular emotion blends.

A common theme in my discussion of blends that support an additive model and blends that have emotions diverging from typical appraisal-emotion relation is the contextual influence. In blends where appraisals from different emotions are combined additively, the contexts in which emotional blends arise may limit what emotions can be realistically blended. In blends where the additive model is contradicted, the blend appraisal profiles may change the function of an emotion through eliciting it without a key appraisal, along with other emotions. From the investigation of appraisal profiles for emotional blends, it seems that appraisals formulate emotional blends only within specific contexts, as if there are rules that determine what emotions can go together and when emotions can function without critical appraisals in particular blends. This observation aligns with existing arguments over the idiosyncratic connections among singular emotions that determine the co-occurrence of emotions within realistic settings (Izard, 1992; Larsen, McGraw, & Cacioppo, 2001). While in this study, I used a data-driven approach to show what emotions co-occur and how appraisals support the co-occurrences, the question over what are the idiosyncratic connections among emotions that constrain the co-occurrence, remains unclear. My interpretation of high anxiety without low congruence in the anxious excitement blend may offer a useful direction: in the absence of key appraisals, it is possible for some other appraisals to partially replace the role of the missing key appraisal, altering the function of an emotion in a way to accommodate other emotions that do not typically cooccur. A systematic exploration along this direction will greatly enhance the knowledge about emotional blends.

A key interest for establishing the emotional blend appraisal theory is to observe how appraisals support various emotional blends, including blends with emotions of opposing valence. The bittersweet blend would make a prime example for understanding appraisals within a mixed or dialectic emotional experience, but disappointingly, a blend resembling the bittersweet blend was not observed from the LPA solution. The lack of a bittersweet blend shows the shortcoming of a data-driven approach for studying emotional blends: with no studies specifically inducing a state of bittersweet included in the dataset, the low rate of occurrence for the bittersweet blend would not be able to form a consistent blend in the LPA solution. What the LPA classes do demonstrate is the existence of blends where opposing valence emotions co-occur, such as the combination of anxiety and happiness in the anxious excitement blend, thus leaving open the possibility of observing a bittersweet blend and investigating its appraisals. In study 2, a focused induction of the bittersweet blend should help to avoid a lack of the bittersweet experience sampled.

Limitations

Study 1 employed a data-driven approach, whose limitation, disregarding what analytic methods were used, could always trace back to the issues associated with the dataset. As the clusters of emotions and the associated appraisals were compared with theoretical predictions based on Table 2, even with a dataset combining many individual studies, some potential emotional blends were inevitably not covered by any of the studies included or not differentiated based on the 9 emotions selected for analysis. This limitation of the data-driven nature led to the lack of a bittersweet blend in the final LPA solution. However, the data-driven approach was still able to recover a majority of the predicted emotions blend, lending support to both the theory and the approach. Furthermore, study 1 found several interesting emotional blends that helped to adjust and expand what was previously predicted for emotional blends. The new discoveries from this study offered insights into building new theories for future empirical investigation.

Another limitation in study 1 was the response scale, which was unalterable given the data-driven nature. While the 1-9 Likert scale provided a high degree of response flexibility, the lack of finer differentiation in response scales could leave some clusters to have virtually no variances on some emotions. However, such worry did not materialize in the model fitting process. To address this limitation, in study 2, the response scale would use finer response levels for self-reported questions, to allow for better differentiation.

CHAPTER 3

Study 2

Study 2 was designed to examine some of the theoretically predicted blends seen in Table 2 as well as those recovered from data in study 1, in a lab setting. The goal is to experimentally induce a couple of very important blends while capturing the induced blends and associated appraisals. While it is conceptually meaningful to study emotional blends as integral experiences that summarize complex environmental implications, emotional blends need to be reliably induced and measured in experiments just like their singular emotion counterparts to yield insights and generate research interest. If anger or other singular emotions were rarely induced and documented, the research on singular emotions would not have flourished, let alone produce systematic findings such as lists of appraisals associated with major emotions. In study 2, emotional blends were induced with directed autobiographical story retelling. The assessment of the blends employed two information channels including self-report and sound analysis, to gather converging evidence. Appraisals were measured through self-reports.

Beyond the contribution of experimentally inducing emotional blends to investigate the emotional experience and the associated appraisals, study 2 was also designed to make a methodological contribution in assessing emotions and emotional blends through machine learning. Experimental measurements of multiple emotions so far have largely been obtained from self-reports, with participants responding to many emotion questions regarding the same emotional experience. For emotional blends research, a tool that measures multiple emotions at the same time can allow participant to focus more on the actual experience of blends, rather than the act of reporting emotions in blends. Emotional information embedded in voice offers a possible route to measure emotional blends in research settings, as the way people speak has been widely documented in past research to convey the speakers' emotional state (Banse & Scherer, 1996; Sorbin & Alpert, 1999; Juslin & Laukka, 2001; Coutinho & Dibben, 2013; Rao, Koolagudi, & Vempada, 2013). Since emotions were induced with retelling, the material produced by the participants provided a good source of emotional information that can complement the subjective reports of emotions.

An advantage of extracting emotional information from sound signals compared to sentiment analysis based on the content of emotional story retelling was the independence from content. Previous emotion decoding studies have used generic sentences produced by speakers with emotional tones, and demonstrated high accuracy in emotion detection (Sorbin & Alpert, 1999). Rather than looking into what participants said in the recordings, sound parameters from the speech included the intensity of sound, frequency of sound, sphericity of sound, and speech rate to serve as the basis for emotion classification. The value of using sound parameters was that in many situations the emotional state was not conveyed directly from text. For example, a statement like "we were great friends back in college" might look pleasant based on text, yet when spoken in a sad tone would convey a deep sense of loss.

To extract emotions from sound, I planned to use voice parameters to predict the emotions within speech. Emotion researchers have successfully documented reliable associations among voice parameters and emotions (Banse & Scherer, 1996; Sobin & Alpert, 1999; Juslin & Laukka, 2001), such as high voice intensity coupled with high fundamental frequency in an angry voice as opposed to low intensity and low fundamental frequency in a sad voice (Banse &

Scherer, 1996; Juslin & Laukka, 2001). Recently, with the popularization of machine learning techniques, research on identifying emotions from voice has gradually begun to receive attention from the computer science field, but has barely made impact on substantive research topics in psychology. In study 2, if there are reliable differences in the voices of participants when retelling events featuring various singular emotions and emotional blends, the utility of machine learning based voice analysis will offer a parallel assessment of multiple emotions for future studies beyond self-report measures.

Method

Participants

A power analysis for study 2 based on a planned repeated measure MANOVA design with a very conservative effect size (1% of total variance accounted for by the difference among repeated measures, Cohen, 1988) and a moderate correlation among repeated measures (.3), required a sample size of 154 for a .80 power at .05 α level. A total of 165 participants were recruited from the Vanderbilt University undergraduate student body and a separate online sample that mostly consisted of undergraduate students. Participants from the Vanderbilt University sample received 2 research credits as a part of their course requirements. Participants in the separate online sample received $10 Amazon gift card. Both samples completed the same experimental procedure that lasted about 1 hour.

Among the 165 participants recruited, 160 participants completed the self-reported measures. 156 participants had complete recordings for the 6 personal emotional stories. The 9 participants without complete recordings either did not complete the study, or completed the study but their data were lost in the later processing stage. Of the 165 participants, there were 124 females. Over all participants, the average age was 21.0 years.

Procedure and Material

Participants were recruited into the study with the knowledge that the experiment looked at examining the properties of emotions through the telling of emotional stories. After a relaxation procedure, the participants were asked to relate a personal experience corresponding to one of the six emotion categories to be induced. The emotion categories included four singular emotions including anger, sadness, happiness, and calmness, as well as two emotional blends including the bittersweet blend and the bitter blend. The bittersweet blend included happiness and sadness, whereas the bitter blend included anger and sadness. The singular emotion stories served as baselines to which the two blends could be compared. The six emotional stories were induced sequentially, with the calm retelling always preceding the other five emotion inductions. The order of the other five inductions were randomized through a Latin Square design across participants. During the four singular emotion inductions, participants were directly instructed to reminisce about personal emotional stories that contained the emotion to be induced. For the two emotional blends, participants were instructed to reminisce about personal emotional events that contain both emotions in the blend. In the instructions, participants were required to write down the gist of the emotional stories that they would retell. Participants had two minutes to write about the emotional event for each induction. This type of prompt helped participants focus on the target emotion(s) during their retelling of the emotional events. Table 6 lists the emotion categories next to their instructions.

After the two minutes of writing had elapsed, an instruction for the autobiographical retelling preparation was shown. The following instructional text was designed for participants to think about a happy experience, with changes to the text made accordingly depending on the emotion or emotions induced:

Thank you for writing down a happy event that recently happened to you. In the next minute, please try to recreate the situation you experienced in your mind as if you were in the situation right now. As you relive the situation, please try to attend to as many details as possible. Please pay attention to what you are thinking and how you are feeling.

The preparation instruction was displayed on a screen in front of the participant, with a timer displayed beside the instruction counting down from 60 seconds. The timer was refreshed every 20 seconds. After the one minute elapsed, the participants were instructed to begin speaking about their emotional stories. To promote consistency in recording quality, participants were instructed to maintain their posture and relative distance to their recording devices as much as they could during the recording sessions.

Table 6. Instruction to select a recent event

Emotion/Blend	Instruction
Calm	Please think about a recent personal experience during which you felt calm.
Angry	Please think about a recent personal experience during which you felt angry.
Sad	Please think about a recent personal experience during which you felt sad.
Happy	Please think about a recent personal experience during which you felt happy.
Bitter blend	Please think about a recent personal experience during which you felt angry and sad at the same time.
Bittersweet blend	Please think about a recent personal experience during which you felt sad and happy at the same time.

Instead of allowing participants to retell their emotional stories however they wanted, a guided retelling paradigm lasting 2 minutes was found to be effective in inducing emotions of only the intended types (Labouvie-Vief, Lumley, Jain, & Heinze, 2003). Labouvie-Vief and colleagues displayed the following questions on a poster in front of participants as a reminder for retell:

What happened? What brought it about? How did you feel? How did it go away?

The questions tapped into the details of past emotional encounters to effectively induce emotions. In another study, Smith and Ellsworth asked participants a series of questions about their thoughts and feelings regarding specific emotional experience, in order to boost the accuracy of subsequent reporting of appraisals (1985). Because of the interest to capture appraisals associated with emotional blends in the current study, the retelling instruction involved guiding participants to answer the following questions using the happiness induction as an example, with most adapted from Smith and Ellsworth (1985):

1. Please describe this past happy situation. What was it like to be in this happy situation?

2. What happened in this happy situation that made you feel happy?

3. Why did things happening in the situation make you feel happy?

4. What did it feel like for you to be happy in this situation?

5. What did you do in this happy situation?

All the questions were shown on the screen together throughout the retelling session. The participants were asked to talk about their emotional experiences according to the questions, but not necessarily in the order shown in the list. Participants were assured that there were no right or wrong answers, nor did the length of answer to each question matter. Next to the instructions, the screen displayed a timer that counted down from 120 seconds to 0, with a refresh at every 20 seconds. The participants were encouraged to speak until the timer ends, but to stop at that point whether or not they answered all the questions. A simulated screen is shown in Figure 9. At the end of the retelling section, a brief questionnaire assessing the participants' current appraisals and emotions was administered for each emotion induction. After the self-report questions, the participants were instructed to count backward from 30 to 1, to reset their emotional state before the next emotional induction (Labouvie-Vief, Lumley, Jain, & Heinze, 2003). The guided retelling for the calm story used the same set of on-screen questions, with the same countdown timer indicating how much time was left in the retelling task.

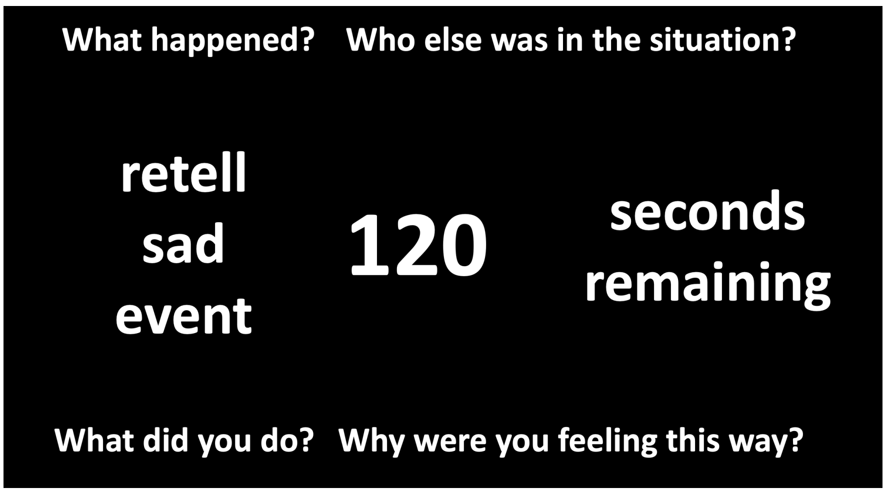


Figure 9. Simulated participant screen

The outbreak of Covid-19 rendered in-person data collection unsafe for researchers and participants, calling for a socially-distanced data collection plan. All participants received a Zoom video conference platform link to join the experiment virtually. All experimental procedures were administered through Zoom. The audio recording component, crucial to study 2, was conducted through the recording functions in Zoom. While video was recorded as a part of the recording procedure, they were discarded as soon as possible after the experiment to protect the privacy of participants.

Measures

Self-Report Measures.

For self-report appraisal assessments after each retelling, the 7 appraisals from Table 2 except for congruence were used. Instead of a single congruence item, participants rated their appraised levels of congruence and incongruence with separate items. Such a split allowed for assessing the congruence appraisal as two unipolar scales rather than a single bipolar scale, to freely evaluate situations like the bittersweet blend where both congruence and incongruence were hypothesized to exist (Larsen & McGraw, 2011). Right after the appraisal measures, the participants reported the same 9 emotions from Table 2. The measurement scale adopted a text box design that asked participants to manually type a numeric value between 1 and 100 for each appraisal and emotion. This response scale was expected to yield more variation than the 1-9 Likert scale. The emotion and appraisal assessment for the calmness induction did not differ from the assessments after other inductions. After all 6 emotion retellings were covered, the participants answered a brief demographic questionnaire to conclude the experimental session.

Voice-Based Emotion Measures.

To measure emotions that were projected to appear in the sound recordings of the 4 singular emotions and the 2 emotional blends, a four class emotion classifier was trained to give a probability to each of anger, sadness, happiness, and calmness for an emotional speech sample. If the classifier predicted only a high probability of anger, with probabilities for the other three emotions low, then the classifier would determine that anger was the predominant emotion in the speech sample. If no emotions were present, the calm emotion probability would be high. If both happy and sad probabilities were high, the emotional speech was determined to convey a blend of the two emotions.

The classifier was trained based on a set of emotional speech samples gathered from openly available academic resources[1]. Three speech datasets that employed 119 actors to announce short sentences with pre-determined emotional states contained many sentence-length emotional speech samples for each of the four emotions to be classified. Each actor provided between 28 and 75 speech samples. In total, there were 1523 speech samples for anger, 1523 samples for happiness, 1303 samples for calmness, and 1522 samples for sadness. On average, each speech sample lasted about 3.9 seconds. For each of the speech sample, I extracted several voice parameters that have been documented to correlate with emotional voice in past research (Banse & Scherer, 1996; Sobin & Alpert, 1999; Juslin & Laukka, 2001), including voice intensity, mean fundamental frequency, median fundamental frequency, standard deviation of fundamental frequency, minimal fundamental frequency, maximal fundamental frequency, unvoiced segments, jitter, shimmer, and harmonics-to-noise ratio. A preprocessing script created in Praat speech processing software was used to extract the sound parameters (Boersma & Weenink, 2020). The speech parameters were used to train the emotion classifier.

In creating the emotion classifier, one important step in processing the speech samples was to standardize samples within each person. Since people spoke differently, an emotion classifier trained on one person might misclassify a jovial speech of another person as anger simply because the other person had a deeper voice signature. To accommodate this attribute of

[1] The training speech samples incorporated the Surrey Audio-Visual Expressed Emotion Database (http://kahlan.eps.surrey.ac.uk/savee/Database.html), the Crowd-sourced Emotional Multimodal Actors Dataset (https://github.com/CheyneyComputerScience/CREMA-D), and the Ryerson Audio-Visual Database of Emotional Speech and Song (https://zenodo.org/record/1188976).

speech, all training speech samples for each of the 119 actors were standardized within each actor. Hence, rather than simply classifying the emotion of a particular speech, the classifier determined the emotional state of a speaker given the overall speech profile of the same person. This approach was shown to yield high accuracy in an emotional sound decoding study (Sorbin & Albert, 1999).

After the standardization of the sound parameters, a support vector machine (SVM) was trained based on all actor speech samples. To guard against over-fitting, the SVM was validated through a 10-fold cross-validation with a random 70-30 split of the actor speech dataset into training and testing data. Across the 10 validations, after the SVM was fitted with training sets, the average accuracy of recovering the true emotion label based on sound parameter inputs was around 71% for the four emotions in the testing sets. A confusion matrix that summarizes the percentage of accuracy and misclassification errors across the 10 validations is shown in Table 7. Overall, the SVM showed good accuracy and generalizability. The SVM trained with the full speech sample was used for measuring emotions in the autobiographical retelling recordings.

Table 7. Cross-validation accuracy of SVM emotion classifier

		Predicted label			
		Anger	Happy	Calm	Sad
True label	Anger	81%	13%	4%	1%
	Happy	19%	61%	15%	5%
	Calm	2%	9%	66%	23%
	Sad	1%	6%	17%	76%

For each of the autobiographical story retelling recordings, 115 segments of sound were extracted, beginning from the first second of the recording with a 5-seconds moving window and a 1 second offset. The 5-seconds moving window made the new sound input to which the trained SVM was applied roughly matched the length of the actors' training speech samples. Each segment was entered into the same preprocessing script used for parameterizing actor speech samples. Each of the sound parameters for all segments of a specific participant was standardized to adjust for individual differences in voice parameters. The pre-trained SVM predicted emotion probabilities based on standardized sound parameters, computing probabilities of four emotions for each segment. To make probabilities comparable, the probabilities for each predicted emotion from all recording segments made by a participant were transformed into percentile ranks. The 115 segments from each emotional story retell recording were then summarized by the median percentile rank for each of the four emotions. This meant that if a participant sounded sad the most in a sad story retelling, this participants would have the highest median percentile rank on predicted sadness for the sad story. In the end, each participant had four summary median percentile ranks, including calmness rank, anger rank, happiness rank, and sadness rank for each of the six retelling recordings.

Analysis Plan

The first part of the analysis used self-reported appraisals and emotions to investigate the viability of experimentally inducing emotional blends, as well as to identify the appraisals associated with induced emotional blends. A repeated-measure MANOVA was conducted to test for overall differences in appraisals and emotions measured through self-report across various emotion categories. Subsequently, post-hoc comparisons with Type I error corrections were conducted to examine how appraisals and emotions differed across the different stories. With the post-hoc comparisons, I conducted a manipulation check to test whether the different stories induced the intended emotions.

The voice-based emotion analysis followed the analysis of self-reported appraisals and emotion. A repeated measure MANOVA was conducted on the median percentile ranks for all four emotions measures from the six emotion inductions. If overall differences were established, post-hoc comparisons for each predicted emotion rank were conducted first among the singular emotion retellings to test if the algorithmic emotion prediction could recover elevation in the emotion induced for each retelling. To examine how voice differed between blend retellings and singular emotion retellings on each predicted emotion, for each blend, a set of post-hoc comparisons was conducted among blend story and the stories of the blend constituting singular emotions. To draw comparisons between the self-reported emotions and the machine learning computed emotions, Pearson correlations were used to examine if each of the predicted emotions had a positive correlation between the self-reported measure and the machine learning prediction for both the overall sample and within each story.

Hypotheses and Predictions

I predicted that across the six emotion induction categories, the repeated measure MANOVA would show significant differences across the self-reported measures of emotions and appraisals. This would mean the emotional blend induction method was effective in inducing different emotional experiences with distinctive cognitive foundations. Post-hoc comparisons would further support the viability of inducing emotional blends experimentally by showing blends and singular emotions differed on the emotions and appraisals. Table 8 lists the emotions and appraisals that were predicted to be associated with each type of story, which were checked with the post-hoc comparisons.

Table 8. Theoretically predicted appraisals and emotions induced in different stories

Emotion/Blend	Relevance	Congruence	Incongruence	Self	Other	Future	PFCP	AFCP	Anger	Guilt	Anxiety	Sadness	Hope	Challenge	Happy	Pride	Gratitude
Calm	Medium	Medium	Medium	/	/	Medium	/	/	/	/	/	/	/	/	/	/	/
Angry	High	Low	High	/	High	/	/	/	High	/	/	/	/	/	/	/	/
Sad	High	Low	High	/	/	Low	Low	/	/	/	/	High	/	/	/	/	/
Happy	High	High	Low	/	/	/	/	/	/	/	/	/	/	/	High	/	/
Bitter blend	High	Low	High	/	High	Low	Low	Low	High	/	High	High	/	/	/	/	/
Bittersweet blend	Medium	High	High	/	/	Low	Low	/	/	/	/	High	/	/	High	/	/

Note. “/” represents no specific prediction is given for the emotion or appraisal.

In the sound analysis, I predicted that across the six emotion induction categories, the repeated measure MANOVA would show significant differences across the median ranks of emotion predictions based on the SVM. Furthermore, post-hoc comparisons would reveal how different emotional retellings differed on each algorithm predicted emotion, as well as how the voices from the blends differed from the voices from the singularly emotion stories. Finally, I expected the SVM predicted emotions to positively correlate with the self-reported emotions in the overall sample and within each story.

Results

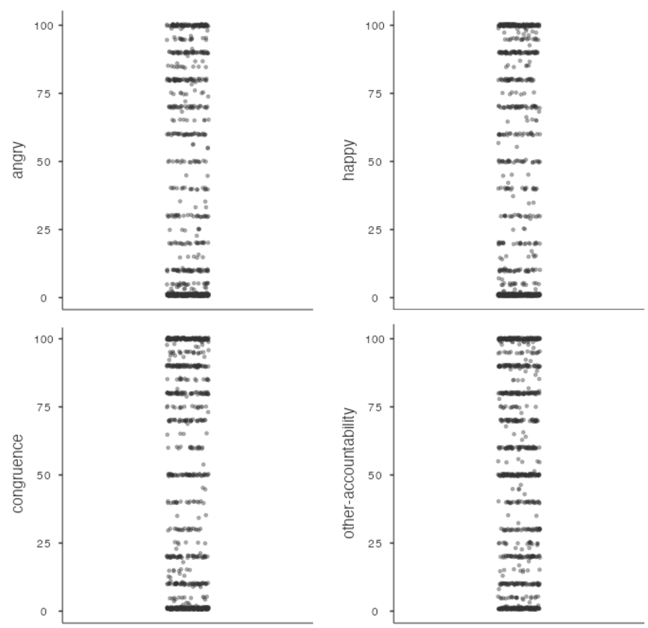


Figure 10. Distributions of a few variables

A quick check over the response patterns revealed that despite given a fine response scale, participants largely treated the scale as a 11 point scale including 1 and 100, with most responses in between occurring at the numbers of whole tens. Figure 10 depicts the responses of all participants on anger, happiness, congruence, and other-accountability to sample how participants used the response scale. The concentration of responses at the numbers of whole tens, as well as at both extremes, is contrasted by the sparsity of responses in the space between the bands of concentrated responses.

Self-Reported Emotions

Figure 11 presents the average self-reported emotions for each of the 6 types of stories. Beyond the emotions that were targeted in the induction prompts, there was at least one other emotion with elevated ratings in the emotion profile for each type of story, such as elevated hope in happy, calm, and bittersweet stories, as well as elevated anxiety in angry, sad, bitter, and bittersweet stories. The elevation in the emotions not targeted in the induction prompts demonstrated that participants had genuine experience of emotional blends, rather than merely endorsing emotions they saw in the prompts.

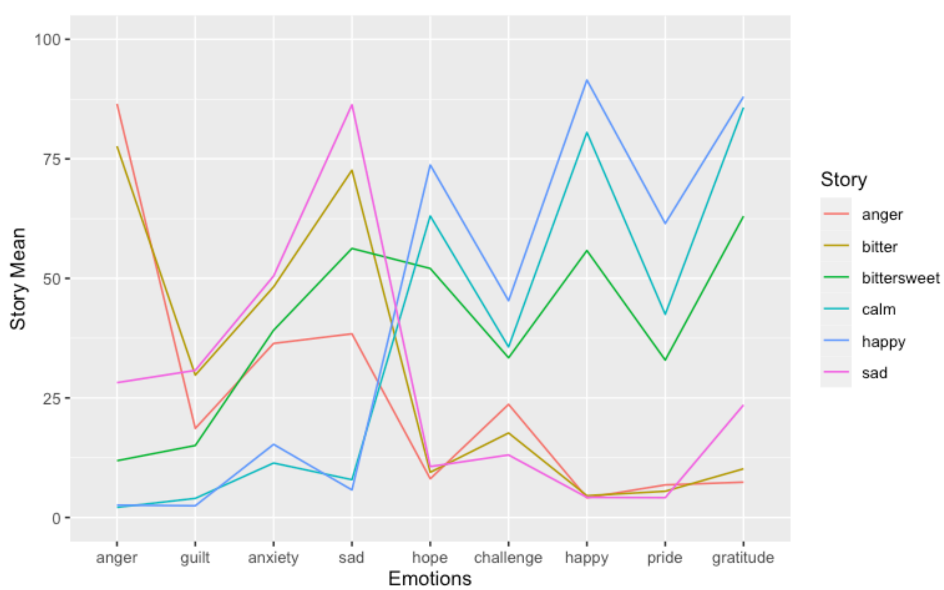


Figure 11. Story average self-reported emotions

A repeated measures MANOVA was conducted to test for the overall difference among emotions across different types of stories. Due to high non-normality in the self-reported emotions, a nonparametric version of repeated measures MANOVA was conducted with the multRM function within the RM.MANOVA package in R (Friedrich, Konietschke, & Pauly, 2021). The modified ANOVA-type statistic was 18344.3, with a p-value of <.001 obtained from wild bootstrapping, indicating a significant overall difference among emotions across the 6 different story retellings. The wild bootstrapping approach, which multiplied a random weight to the centered bootstrap sample, and the set of statistics were recommended by the authors of the repeated measures MANOVA function adopted for this analysis (Friedrich & Pauly, 2018). ,

The post-hoc comparisons, conducted based on a series of univariate repeated measures ANOVA (jamovi, 2021), helped to test if emotion inductions were successful. For each emotion, all pairwise post-hoc comparisons with Tukey correction, along with the table of the estimated marginal means for each type of story, are reported in Supplementary Table 1. To test for emotion induction efficacy, the predicted differences between stories from Table 7 were examined against the post-hoc comparisons, with the supported predictions updated with observed ones in the parentheses in Table 9. The levels designated in the cells were based on the average rating for each emotion. In each emotion, the percentile rank of story averages along the spread between the highest story average rating and the lowest story average rating categorized each type of story as low (0th-33rd percentile), medium (34th-66th percentile), or high (67th-100th percentile), on that emotion. The emotions that were predicted to elevate from story retelling were successfully induced, except that sadness and happiness were somewhat less elevated in the bittersweet story than expected. The calm story retelling presented an emotion profile that did not significantly differ from the happiness story on anger, guilt, anxiety, sadness, and gratitude (all $p_{tukey} > .769$), while significantly different from but nevertheless closely tracking the happiness story on hope, challenge, happiness, and pride (all $p_{tukey} < .017$).

Table 9. Post-hoc comparison summary for self-reported emotions

Emotion/Blend	Anger	Guilt	Anxiety	Sadness	Hope	Challenge	Happy	Pride	Gratitude
Calm	/	/	/	/	/	/	/	/	/
Angry	High	/	/	/	/	/	/	/	/
Sad	/	/	/	High	/	/	/	/	/
Happy	/	/	/	/	/	/	High	/	/
Bitter blend	High	/	High	High	/	/	/	/	/
Bittersweet blend	/	/	/	High (Medium)	/	/	High (Medium)	/	/

Note. Observations described in the parentheses are those that deviated from the prediction from Table 7. Low, high, medium designation for each story is based on where the story average occupies in the three part even division of the space between the highest and lowest story averages for each appraisal.

Further investigation of the self-reported emotions, based on both the story average seen in Figure 12 and the post-hoc comparison results in Supplementary Table 1, helped to inform how the emotional blends differed from the constituting singular emotions across the whole range of emotions assessed. For the bitter story, the emotion profile was largely nested between the anger story profile and the sadness story profile, significantly different from the two stories on all negative emotions (all $p_{tukey} < .001$), except for guilt and anxiety when compared to the sadness story (both $p_{tukey} > .973$). In terms of mean differences, the bitter story emotion profile was much closer to the anger story on anger ratings ($D = 8.869$) than the sadness story ($D = 49.450$), as well as much closer to the sadness story on sadness ratings ($D = 13.687$) than the anger story ($D = 34.213$). The bittersweet story emotion profile was nested in between the sadness story emotion profile and the happiness story emotion profile, with the bittersweet emotion profile significantly different from the story profiles of the two constituting emotions on all the self-reported emotions (all $p_{tukey} < .002$). Although lower in overall intensity, the shape of the bittersweet blend emotion profile matched the sadness story on the emotion profile on anger, guilt, anxiety, and sadness, whereas the blend matched the shape of the happiness story profile on hope, challenge, happiness, pride, and gratitude. Compared to the bitter emotion profile, the bittersweet emotion profile had a significantly lower set of negative emotions and a significant higher set of positive emotions (all $p_{tukey} < .025$).

Self-Reported Appraisals

Figure 12 presents the average self-reported appraisals for each of the 6 types of stories.

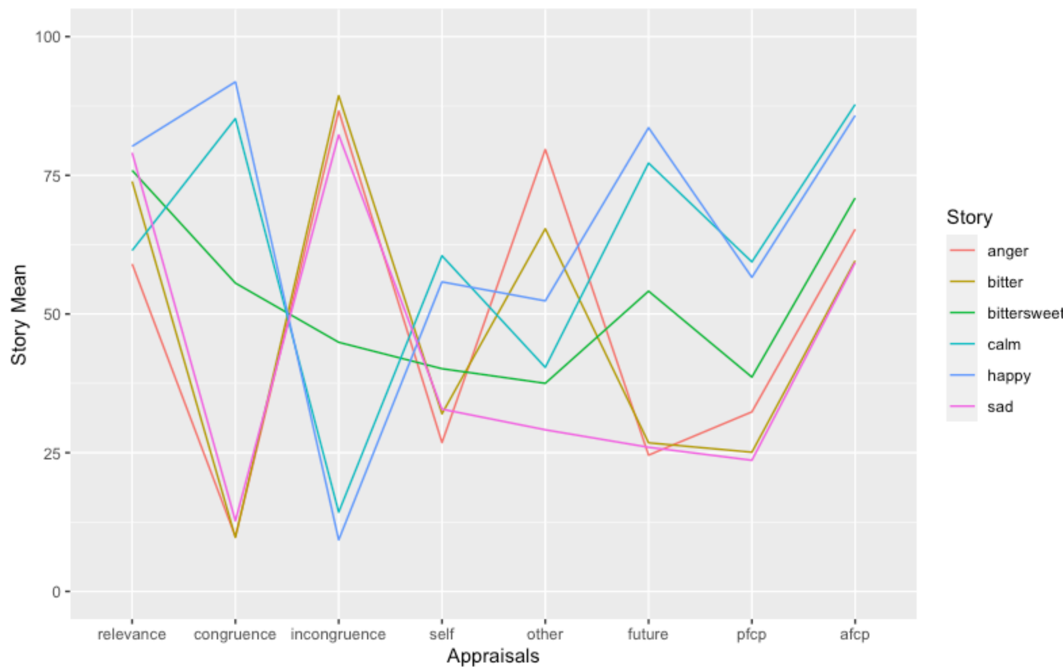


Figure 12. Story average self-reported appraisals

To test for the overall difference in self-reported appraisals across different stories, I conducted a repeated measure MANOVA over the 8 self-reported appraisals. The same non-parametric version used in the self-reported emotion analysis was used again here. The modified ANOVA-type statistic is 9153.8, with a p-value of <.001 obtained from wild bootstrapping, indicating a significant overall difference in appraisals across stories.

For each appraisal, all pairwise post-hoc comparisons with Tukey correction, along with the table of the estimated marginal means for each type of story, are reported in Supplementary

Table 2. The predicted differences from Table 7 were examined with respect to the post-hoc comparisons, with the unsupported predictions updated in parentheses in Table 10. The process of levels designation in the cells followed the same process used for the self-reported emotions. For the singular emotions, the story retelling was associated with most of the predicted appraisals, except that the ratings for congruence, incongruence, and future expectancy were more positive than predicted for the calm story. Additionally, the angry story had lower relevance appraisals than expected, although the self-reported anger in this story was high.

Table 10. Post-hoc comparison summary for self-reported appraisals

Emotion/Blend	Relevance	Congruence	Incongruence	Self	Other	Future	PFCP	AFCP
Calm	Medium	Medium (High)	Medium (Low)	/	/	Medium (High)	/	/
Angry	High (Medium)	Low	High	/	High	/	/	/
Sad	High	Low	High	/	/	Low	Low	/
Happy	High	High	Low	/	/	/	/	/
Bitter blend	High	Low	High	/	High	Low	Low	Low
Bittersweet blend	Medium (High)	High (Medium)	High (Medium)	/	/	Low (Medium)	Medium	/

Note. Observations described in the parentheses are those that deviated from the prediction from Table 7. Low, high, medium designation for each story is based on where the story average occupies in the three part even division of the space between the highest and lowest story averages for each appraisal.

Based on both the story average seen in Figure 13 and the post-hoc comparison results in Supplementary Table 2, I made following observations about the differences in the key appraisals between the two blends and their constituting emotions. Overall, the appraisals of the two blends were nested in between the appraisals of their respective constituting emotions, where the constituting emotions differ. Specifically, compared to the angry story, the bitter blend had a significantly lower other-accountability ($p_{Tukey} < .001$), but did not differ on congruence and incongruence (both $p_{Tukey} > .865$). Because of the unexpected low relevance for the angry story, the bitter blend had a significantly higher relevance rating than the angry story ($p_{Tukey} < .001$). Compared to the sad story, the bitter blend did not differ on congruence ($p_{Tukey} = 0.771$) or PFCP

(p_{Tukey} = .997), but had a significantly higher incongruence appraisal (p_{Tukey} = .044). The bittersweet blend differed from the sad story with a significantly higher congruence, a significantly lower incongruence, and a significantly higher PFCP (all p_{Tukey} < .001). Compared to the happy story, the bittersweet blend had a significantly lower congruence and a significantly higher incongruence (both p_{Tukey} < .001). Contrary to a medium relevance prediction, the bittersweet blend did not differ from either the sad or the happy strong on relevance (both p_{Tukey} > .602).

Voice-Based Emotion Detection

Based on the predicted emotion with the top predicted probability received for each voice segment, around 39% of all voice segments obtained in this study were deemed happy, in addition to around 27% angry, around 21% sad, and around 13% calm. A repeated measures MANOVA over all 4 predicted emotion percentile rank variables found a significant overall difference across the 6 types of emotional stories, the modified ANOVA-type statistic is 138.7, with a p-value of <.001 obtained from wild bootstrapping.

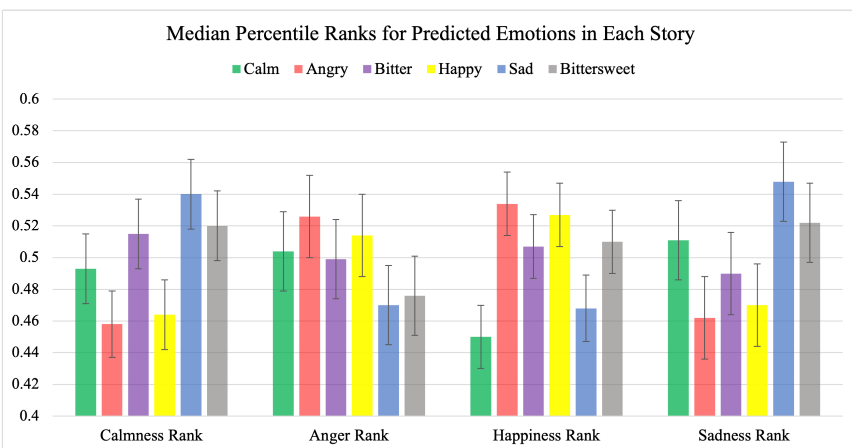


Figure 13. Average median percentile ranks for predicted emotions in each story

Figure 13 presents the story average median percentile ranks on each predicted emotion. The error bars represent the 95% confidence interval of the estimated marginal means. The estimated marginal means, presented in Supplementary Table 3, were obtained from a series of univariate repeated measures ANOVA.

All pairwise comparisons among the singular emotion stories on each of the 4 predicted emotions are presented in Supplementary Table 4. For predicted calmness rank, the calm story was significantly lower than the sad story ($p_{Tukey} = .026$). The sad story was also significantly higher than the happy and the angry story (both $p_{Tukey} < .001$). For predicted anger rank, none of the singular emotion stories were significantly different from each other (all $p_{tukey} > .057$). For predicted happiness rank, the happy story was significantly higher than the calm and the sad story (both $p_{Tukey} \leq .001$). However, the happy story was not different from the angry story ($p_{Tukey} = .970$), which was also significantly higher than the calm and the sad story (both $p_{Tukey} < .001$). Finally, for predicted sadness rank, the sad story was significantly higher than the angry and the happy story (both $p_{Tukey} \leq .001$).

Post-hoc comparisons between each blend story and the stories of the blend constituting emotions are presented in Supplement Table 5. The bitter story was not different from the angry and the sad story on the predicted anger rank (both $p_{Tukey} > .249$). The bitter story was significantly lower than the sad story on the predicted sadness rank ($p_{Tukey} = .001$). For the bittersweet story, it was significantly higher than the happy story on the predicted sadness rank ($p_{Tukey} = .017$), and significantly higher than the sad story on the predicted happiness rank ($p_{Tukey} = .023$).

Overall for the singular emotion stories, the voice analysis could not reliably differentiate angry stories from happy stories, as was evident in the predicted anger and happiness ranks for these two stories. Calm and sadness stories also presented challenges to be differentiated reliably on their voice signature. The voice analysis was able to tell apart sad stories from happy or angry stories, as was shown in the predicted happiness and sadness ranks. For blend stories, the bittersweet story showed considerable blending of happy and sad voices compared to both the happy and the sad stories, although the bitter story voice was not reliably different from the angry story voice on the predicted sadness and anger ranks.

Method Correspondences

For the overall sample aggregated across different emotional stories, the Pearson correlations for the 4 predicted emotion median ranks with the corresponding self-reported emotions were low, as seen in the overall column in Table 11. Except for the sadness prediction, where there was a significant and positive correlation, the other three correlations were not significantly different from 0. With the data broken-down by different types of stories, Pearson correlations between the two measures on all predicted emotions for all types of stories were not significantly different from 0.

Table 11. Pearson correlations between self-reported emotions and predicted emotion ranks

Emotion measures	Overall	Breakdown by Story					
		Calm	Angry	Bitter	Happy	Sad	Bittersweet
Calm	-0.031 (-0.094, 0.033)	0.046 (-0.111, 0.202)	0.041 (-0.117, 0.196)	0.036 (-0.121, 0.192)	0.057 (-0.100, 0.212)	0.095 (-0.062, 0.248)	0.025 (-0.132, 0.181)
Angry	0.011 (-0.053, 0.074)	0.145 (-0.011, 0.295)	-0.037 (-0.192, 0.120)	0.038 (-0.12, 0.193)	-0.18 (-0.327, -0.024)	-0.088 (-0.242, 0.069)	-0.059 (-0.213, 0.099)
Happy	0.005 (-0.059, 0.069)	0.069 (-0.088, 0.224)	0.094 (-0.063, 0.247)	-0.032 (-0.188, 0.125)	0.069 (-0.088, 0.224)	-0.014 (-0.170, 0.143)	0.118 (-0.040, 0.269)
Sad	0.112 (0.048, 0.175)	0.025 (-0.132, 0.181)	0.08 (-0.078, 0.234)	-0.017 (-0.173, 0.140)	0.11 (-0.048, 0.262)	-0.005 (-0.161, 0.152)	0.072 (-0.085, 0.226)

Note. Self-reported emotions for each participant were paired with the median percentile ranks for the corresponding emotions from the voice analysis. The single number in each cell is the correlation point estimate. The parentheses contain the 95% confidence interval for the correlation estimates.

Discussion

Study 2 was designed to experimentally induce two specific emotional blends, a goal that was met successfully with the autobiographical story retelling. The two emotional blends saw their unique set of constituting emotions elevated in the self-reported emotion measures relative to other singular emotion stories. The distinctiveness of the emotional blends was further demonstrated by the recovery of the predicted appraisal profiles from self-report measures. In an largely additive fashion, the two blends retained the key appraisals for all of their constituting singular emotions, such as elevated other-accountability in the bitter blend similar to anger, and the depressed PFCP in the bittersweet blend similar to sadness. Together, the analyses of self-reported emotions and appraisals in study 2 supported the emotional blends as specific emotional entities with characteristic emotional experiences and appraisal foundations.

The singular emotion stories had a wide variety of elevated emotions, even though participants were given a single emotion as prompt. The rich emotional experience from the recalled event demonstrated the efficacy of the autobiographical retelling design in inducing participants into the organic emotional encounter, rather than merely leading them to give high scores to the emotions in the prompt. Furthermore, the emotion profiles of the singular emotions corroborated previous research findings on the prevalence of co-occurring emotions in a wide variety of emotional encounters (Ellsworth & Smith, 1987; Robinson & Clore, 2001).

The distinctiveness in self-reported emotions and appraisals of the emotional blends also established the feasibility of inducing emotional blends with a retrospective recalling approach. While the autobiographical retelling design explicitly asked participants to think about specific sets of emotions, the resulting self-report emotion measures demonstrated marked difference on not only the emotions that were mentioned in the prompts, but also other emotions associated with some of the constituting emotion but never explicitly mentioned, such as the elevated anxiety for the bitter blend and the elevated hope and gratitude for the bittersweet blend. The rich emotional experience that participants reported lent support to the format of emotional blend inductions used in this study.

The self-reported emotions and appraisals offered an in-depth look into the relation between the blends and the singular emotion stories. For the bitter blend, in comparison to the angry story and the sad story, its emotion profile contains elevated anger and sadness similar to the two singular emotion stories but with less intensity. Due to a lack of differences, the congruence and incongruence appraisals could not explain the significantly lower reported anger in the bitter story compared to the angry story, but a significantly lower other-accountability appraisal reported in the bitter story than the angry story could. The medium level of other-accountability in the bitter blend supports the additive model in averaging the low level of other-accountability in the sad story and the high level in the angry story. The lower reported sadness in the bitter story compared to the sad story could not be explained with any significant appraisal differences. Although highly speculative, a crowd-out explanation could be that a higher other-accountability elevated the anger component in the blend, which in turn limited the sadness component.

The blending of multiple emotions but with less intensity was also true for the bittersweet blend. With less intense happy and sad emotion ratings, the bittersweet blend had a combination of medium congruence and incongruence appraisals. For other emotions, the bittersweet blend also had significant elevations, such as a medium level of hope following that of the happy story, and the high level of anxiety following that of the sad story. The simultaneous elevations in self-reported happiness and sadness contradicted the argument for mutual exclusivity between opposite valence emotions (Russell & Carroll, 1999). Instead, the bittersweet blend results supported what Cacioppo and colleagues had argued: at least in situations like a bittersweet story, positive and negative emotions could be freely configured, rather than being mutually exclusive as in singularly positive or negative situations (Cacioppo & Berntson, 1994; Larsen, McGraw, & Cacioppo, 2001). Like the bitter blend, the blending of the two opposing valence emotions in the bittersweet blend followed an additive rule: the averaging of opposing congruence and incongruence levels in the two singular emotions resulted in the medium levels on congruence and incongruence in the bittersweet blend, whereas the high relevance appraisals for anger and happiness were retained in the blend. The bittersweet blend seemed to be a departure from the anxious excitement blend from study 1, where the co-occurring opposing valence emotions saw the appraisal structure for a blend constituting emotion altered.

Contrary to predictions, the bittersweet blend featured high relevance and medium levels of both congruence and incongruence, rather than a medium relevance combined with high levels of congruence and incongruence. This observation implies that a bittersweet experience is more likely to arise in situations that are immediately relevant, although the loss and success aspects of the environment do not hold extremely positive or negative implications. In a situation like graduation, people should directly perceive both the success in finishing the degree and the loss in leaving their familiar environment, but not highly conducive or inconducive to what they do after graduation. As the result of this appraisal pattern, the medium levels of happiness and sadness are likely pointing people to attend to both the success and the loss, but not in an urgent manner that requires immediate intervention.

The voice analysis was able to reveal some systematic differences across different types of story retelling, but the resolution of emotion prediction was considerably poorer than expected. The most reliable difference between stories was between sad retelling voices and happy or angry retelling voices, whereas happy and angry voices were hardly distinguishable based on the predicted emotions. The emotion prediction algorithm might have relied largely on the intensity of voice to tell apart emotions, as both happy and angry voices have been previously documented to feature high voice intensity while a sad voice has low intensity (Banse & Scherer, 1996; Sobin & Alpert, 1999; Juslin & Laukka, 2001). The confusion between happy and angry voices was not confined to study 2 voice data: the SVM emotion prediction algorithm also had the highest rate of misclassification between anger and happy actor-enacted training recordings. Furthermore, the correspondence between voice analysis and self-report emotion measures was largely insignificant except for sadness, which was associated with a very characteristic low-intensity voice. Yet, even for sadness, the significant positive correlation between the two methods was around a small effect of .1.

There were several possible explanations for the poor performance of the voice analysis. First, the experiment was conducted through Zoom video conferencing, rather than a controlled lab environment. The variability in internet connectivity directly influenced the recording quality for each participants. When there were small parts of retellings that were not audible, the experimenters could understand what the participants said with wholistic processing of information and interpolation, which were remedies not available to the emotion prediction algorithm. The distortions in recording would directly impact the coding of sound parameters, which in turn affected the prediction accuracy. Furthermore, vocalization behaviors specific to videoconferencing, such as increased voice intensity over videoconferencing relative to in-person speech (Croes, Antheunis, Schouten, & Krahmer, 2019), could contaminate the expression of emotions through voice. Second, in the speech sample, participants appeared to sound overly happy, which could skew the recovery of predicted emotions in the different stories. Nearly 40% of all voice segments had happy as the emotion label with the highest predicted probability out of the four possible emotion labels. With only one story out of six specifically targeting happy memories, the proportion of voice segments predicted to be happy was still high even if all bittersweet story voice samples were predicted as happy. Although the analyses were conducted with median percentile rankings, which were designed to remove the overall effect of happiness on each predicted emotion, the overly happy voice still compressed the range of variability for each predicted emotion, thus restricting the possible between-story differences on the predicted emotions. The positivity bias in emotional experience could be a reason for the overly happy voice: in this retrospection-based retelling task, the participants might be more likely to remember and communicate the positive elements from past events. The participants could also have spoken under the influence of social desirability, even if inadvertently. This would mean that disregarding the emotional experience of the past events, some participants could have told the story in a largely positive voice to appear more agreeable to the experimenters.

An additional reason for the lack of method convergence in assessing emotions is the mismatch between the speech data used for training the SVM emotion predictor and the participant data to which the emotion predictor was applied. In the training sample, professional actors were instructed to vocalize a predefined sentence with a posed emotion in voice. The uneven quality by actors in posed emotional vocalization could present challenges to the encoding and decoding of emotional information in voice (Scherer, Banse, Wallbott, & Goldbeck, 1991), although the large number of actors used in study 2 should reduce the impact on potential between-actor differences in vocalization. What a large number of actors could not help to avoid though, was the systematic difference between the posed emotional vocalizations and emotional vocalizations within naturalistic settings, a difference that has gradually come to the attention of emotion and machine-learning researchers (Atias & Aviezer, 2020). High quality naturalistic emotion speech training datasets have yet to be made openly available, although emotional nonverbal utterance datasets have been created from naturalistic vocalization materials like YouTube videos (Cowen, Elfenbein, Laukka, & Keltner, 2019). With better training data availability, future automatic emotion prediction algorithms should become better at detecting emotions from voice.

The voice analysis in study 2 was nevertheless able to contribute to the study of emotional blends. The bittersweet story retelling was found to sound significantly happier than the sad story, and significantly sadder than the happy story. Essentially this result says the voices from different emotions can be merged into a blended voice. On one hand, such an observation validates the uniqueness of emotional blends, adding vocal signature to blend-specific characteristics including emotion profiles and appraisal profiles. On the other hand, the successful detection of blended voice in the bittersweet story proves the feasibility of detecting multiple emotions simultaneously through voice, although much work will be needed.

Furthermore, the observation of mixed emotions in voices strengthens the observation from self-report emotions and appraisals: while self-reports might be challenged as conscious responses to fit story retelling prompts, the blending of voice lacked any explicit instruction, thus adding validity to the self-report findings. Still, as exciting as the preliminary evidence for blending of opposite valence emotions in voice, the multiple challenges presented to the emotion prediction from voice require cautious interpretation and generalization of the results.

Limitations

The design of study 2, while successful in inducing separate singular emotions and emotional blends, could be improved in multiple ways. First, the response scale intended to allow for finer levels of differentiation compared to the 9 point Likert scale used in the study 1 dataset, was not successful in encouraging a more continuous distribution of responses. As seen in the scatterplot within Figure 11, the participants treated the 1 to 100 free response scale largely as a 11 point scale including 1, 100, and all whole-ten numbers in between. While self-reported appraisals and emotions were able to reveal highly differentiated appraisals and emotions on a story-average level, the lack of finer differentiation was still disappointing because of potential limitations in the statistical methods applicable to this data. In future studies, instead of a fine response scale, participants should be given Likert-type response scales that can be properly and efficiently utilized.

Second, the SVM emotion prediction was limited by both low audio quality and a mismatch between the characteristic of the training audio sample and the participant voice sample. As mentioned in the discussion, the prediction algorithm was only capable of differentiating emotions largely along levels of voice intensity, which corresponded with the level of activation in emotions. The good performance in detecting finer differentiations between emotions within training actor samples, such as distinguishing between happy and angry voices, was untransferable to the naturalistic participant voice. Future experiments that rely on capturing emotions within voice should consider employing training voice samples that are obtained in a similar condition as the voice samples to which the emotion detection algorithm is applied. Alternatively, the training voice sample can be masked with noise patterns similar to those seen in typical experiment recordings, in order to facilitate a matching in the operation conditions for the algorithm.

Last but not least, study 2 recruited a mostly college-age sample in the U.S.. Given the influence of individual differences on how appraisals are organized, the induction prompts might induce emotional blends and corresponding appraisals very differently across participants from various cultures or age groups. A previous cross-cultural study has shown that during online communications, English speakers from non-Western countries like Singapore are more likely to mix positive and negative emotions within the same sentence than Western countries like Canada (Grossmann, Huynh, & Ellsworth, 2016). Similar findings regarding how emotions associate differently across people have been extended to aging, with people in an older age becoming more prone to simultaneously experience opposing valence emotions than younger people (Ersner-Carstensen et al., 2011; Charles, Piazza, & Urban, 2017). Future studies on emotional blends should consider expanding the participant demographics. The broadening of demographics will also benefit the study of voice-based emotion prediction: a training sample and an experimental sample that both contain a wide variety of speakers will increase the generalizability of study findings.

CHAPTER 4

General Discussion

In a set of two studies, I attempted to study emotional blends from an appraisal theory perspective. In study 1, a clustering structure with high stability indicated the recovered emotional blends had strong signals in data supporting their existence. The appraisals associated with the recovered blends were able to explain the blended emotions, largely following the singular emotion appraisal-emotion relation in an additive fashion, although exceptions existed. In study 2, an autobiographical retelling task was able to successfully induce two emotional blends, which saw the blending of their constituting emotions on self-reported emotions and key appraisals, as well as the blending of happy and sad voices when the participants retold bittersweet stories. Taken together, the results of these studies indicate that blends represent concrete units of emotional experience that are observable and inducible, each with a characteristic profile of appraisals that can explain its existence.

Both studies were able to identify some mechanisms by which the appraisals supported the emotions in a given blend. In most blends, the appraisal-emotion relations largely followed those proposed for singular emotion settings. These observations implies an additive model largely responsible for blending the appraisals from each emotion into a blend. However, appraisals from different emotions cannot be freely combined additively regardless of context, and it is not clear what situational attributes lead to such constraints. There were several contradictions to the additive model, such as the anxious excitement blend seen in study 1, where anxiety was not longer associated with a theoretically predicted low congruence. For these cases that did not conform to an additive view, I offered an explanation based on a possible functional shifts of these emotions as their key appraisals changed. The functional changes for an emotion may be explained with subtle environmental implications captured by appraisals other than those typically associated with that emotion. The non-additive blending of appraisals within emotional blends shows the functional flexibility of certain emotions when they arise within particular blend contexts, with some emotions like anxiety and hope that are elevated in many blends from study 1, becoming more general purpose beyond their roles defined within the singular emotion appraisal theory. Still, much remains unknown about the situational attributes that lead to the non-additivity: while it was tempting to theorize that non-additivity arises in blends with opposing valence emotions, such as in the anxious excitement blend, the bittersweet blend induced in study 2 offered an counterexample.

Despite that the two studies documented the appraisals associated with major emotional blends, future research effort is needed for investigating the environmental attributes that determine appraisals from what emotions can be combined additively and when does the additive model break down. This is certainly a difficult task, as the number of relevant attributes can be very high. One potential situational attribute, which has long been a part of appraisal theory, is the checks on the motivational urges related to emotions. In a high relevance, anger-inducing situation, people may not choose to attack the responsible party out of anger, because of concerns for social appropriateness (Smith & Lazarus, 1990). From an emotional blend perspective, anxiety, which may serve as a cautious check to the motivational urge of anger, arises with anger, as seen in the anger blend from study 1. In contrast, the anger-only class from study 1 had a much lower relevance appraisal than the anger blend, along with a much lower anxiety level. The low relevance of the anger-only class perhaps removed the need for a cautious check on the urge to act out of intense anger, thus reducing the experienced anxiety. While this interpretation of the results is highly speculative, it offers a potential direction to study the situational attributes along the appraisal processes that lead to multiple emotions.

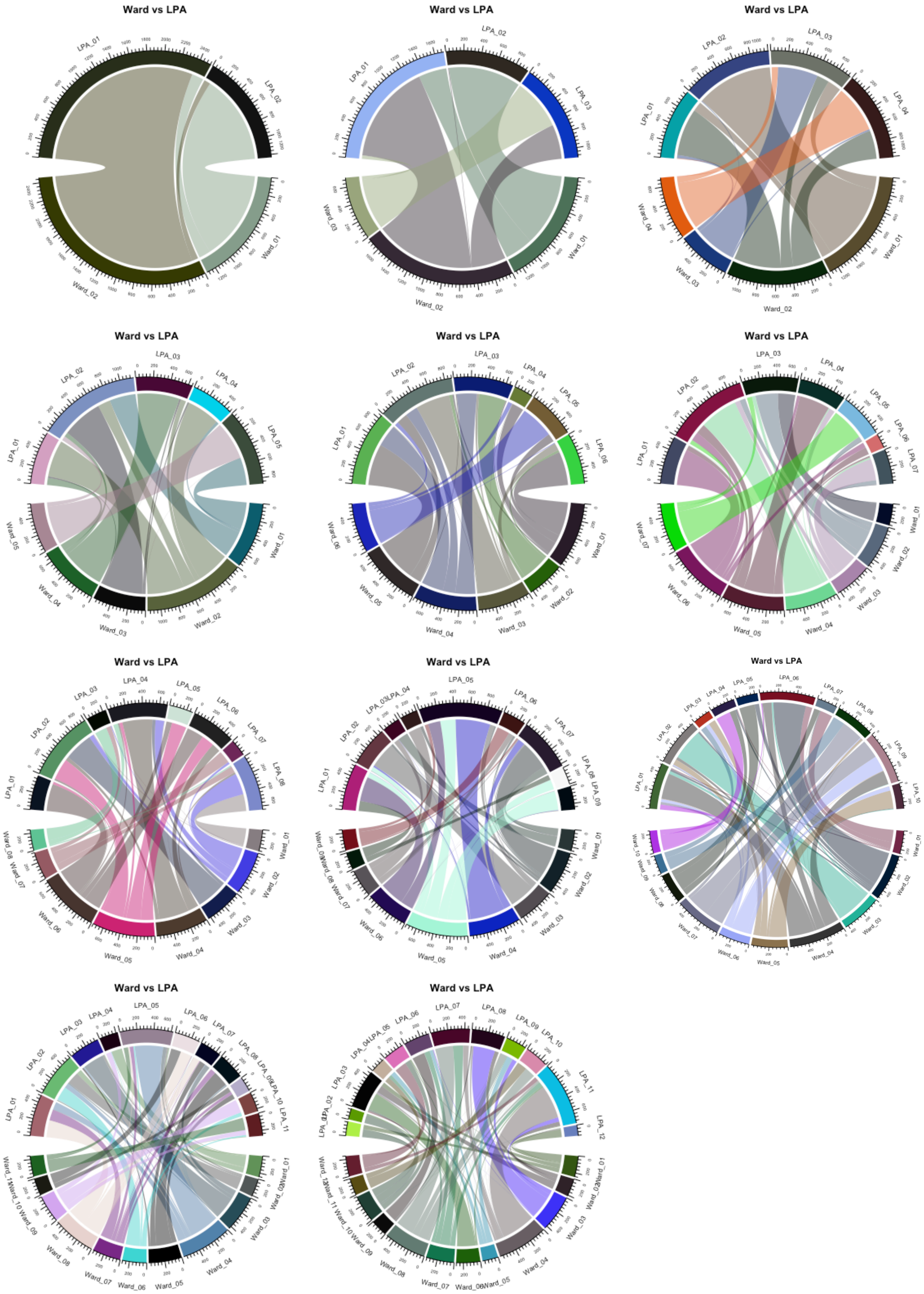
Besides a systematic investigation into the situational attributes that define the appraisal structure within major emotional blends, many aspects of emotional blends are completely unexplored. Different from the cross-sectional view to emotional blends which was taken in both of the current studies, the investigation of temporal dynamics in emotional blends will be able to answer a completely new set of questions including whether multiple emotions in a given blend ebb and flow at around the same time, as well as whether emotional blends can affect ensuing emotional states through cross-temporal facilitation and inhibition effects. Another major aspect of emotional blends worthy of investigation involves individual differences. Questions related to individual differences include whether there are interpersonal level variability in the number of emotions that can be blended, and whether there are personal differences in what situations do people experience blended emotion most. Investigation into these questions may help uncover ties between emotional blends and constructs like emotional intelligence and personal well-being.

Future emotional blend research will likely require efficient assessment of many emotions at once, rather than asking participants to respond to a long list of emotion items each time. The need for an alternative emotion assessment other than self-reports will be especially relevant to temporal dynamic research and individual difference research. Both types of research will likely require high frequency emotion measurements or even continuous emotion monitoring. A voice based assessment, an early implementation of which was able to capture the blending of happy and sad voices, is promising for future use to capture emotional blends in substantive research settings. Although the current version of the emotion detection algorithm was found to be lacking in accuracy, it was nevertheless able to capture emotion activation even in noisy recording environments. More representative training samples that capture emotional vocalization within naturalistic settings will be able to boost the accuracy of detection algorithms in future studies. Other machine learning approaches, such as applying deep learning models to spectrograms of emotional speech samples, might improve prediction accuracy by unearthing extra vocalization information beyond the voice parameters used in study 2.

While far from a comprehensive take on the vast topic of emotional blends, this thesis made an attempt to unpack this previously understudied topic through extracting commonly observed emotional blends, documenting associated appraisal profiles, observing different modes of appraisal-emotion organization within blends, examining the inducibility of blends and associated appraisals, and testing an alternative emotion measurement designed to capture multiple emotions at the same time. Taken together, this appraisal-theory driven attempt to study emotional blends has pushed the understanding of the blends further. Hopefully, with the insight found and the new questions proposed in this thesis, emotional blends can receive increased research interest moving forward.

APPENDICES

Supplementary Figure 1. Chord diagrams for 2-12 class solutions



Supplementary Table 1. Post hoc comparisons for self-reported emotions in study 2

1.1.1 Anger post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-84.394	2.02	795	-41.771	< .001
	-	Bitter	-75.525	2.02	795	-37.381	< .001
	-	Happy	-0.463	2.02	795	-0.229	1.000
	-	Sad	-26.075	2.02	795	-12.906	< .001
	-	Bittersweet	-9.750	2.02	795	-4.826	< .001
Angry	-	Bitter	8.869	2.02	795	4.390	< .001
	-	Happy	83.931	2.02	795	41.542	< .001
	-	Sad	58.319	2.02	795	28.865	< .001
	-	Bittersweet	74.644	2.02	795	36.945	< .001
Bitter	-	Happy	75.063	2.02	795	37.152	< .001
	-	Sad	49.450	2.02	795	24.475	< .001
	-	Bittersweet	65.775	2.02	795	32.556	< .001
Happy	-	Sad	-25.612	2.02	795	-12.677	< .001
	-	Bittersweet	-9.287	2.02	795	-4.597	< .001
Sad	-	Bittersweet	16.325	2.02	795	8.080	< .001

1.1.2 Anger estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	2.12	1.51	-0.845	5.08
Angry	86.51	1.51	83.549	89.48
Bitter	77.64	1.51	74.680	80.61
Happy	2.58	1.51	-0.383	5.55
Sad	28.19	1.51	25.230	31.16
Bittersweet	11.87	1.51	8.905	14.83

1.2.1 Guilt post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-14.631	2.58	795	-5.664	< .001
	-	Bitter	-25.781	2.58	795	-9.981	< .001
	-	Happy	1.544	2.58	795	0.598	0.991
	-	Sad	-26.775	2.58	795	-10.366	< .001
	-	Bittersweet	-11.063	2.58	795	-4.283	< .001
Angry	-	Bitter	-11.150	2.58	795	-4.317	< .001
	-	Happy	16.175	2.58	795	6.262	< .001
	-	Sad	-12.144	2.58	795	-4.701	< .001
	-	Bittersweet	3.569	2.58	795	1.382	0.738
Bitter	-	Happy	27.325	2.58	795	10.578	< .001
	-	Sad	-0.994	2.58	795	-0.385	0.999
	-	Bittersweet	14.719	2.58	795	5.698	< .001
Happy	-	Sad	-28.319	2.58	795	-10.963	< .001
	-	Bittersweet	-12.606	2.58	795	-4.880	< .001
Sad	-	Bittersweet	15.713	2.58	795	6.083	< .001

1.2.2 Guilt estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	3.99	1.92	0.222	7.77
Angry	18.63	1.92	14.853	22.40
Bitter	29.78	1.92	26.003	33.55
Happy	2.45	1.92	-1.322	6.22
Sad	30.77	1.92	26.997	34.54
Bittersweet	15.06	1.92	11.284	18.83

1.3.1 Anxiety post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-25.01	2.94	795	-8.507	< .001
	-	Bitter	-36.84	2.94	795	-12.531	< .001
	-	Happy	-3.91	2.94	795	-1.329	0.769
	-	Sad	-39.11	2.94	795	-13.300	< .001
	-	Bittersweet	-27.77	2.94	795	-9.444	< .001
Angry	-	Bitter	-11.83	2.94	795	-4.024	< .001
	-	Happy	21.11	2.94	795	7.178	< .001
	-	Sad	-14.09	2.94	795	-4.793	< .001
	-	Bittersweet	-2.76	2.94	795	-0.937	0.937
Bitter	-	Happy	32.94	2.94	795	11.202	< .001
	-	Sad	-2.26	2.94	795	-0.769	0.973
	-	Bittersweet	9.08	2.94	795	3.086	0.025
Happy	-	Sad	-35.20	2.94	795	-11.972	< .001
	-	Bittersweet	-23.86	2.94	795	-8.116	< .001
Sad	-	Bittersweet	11.34	2.94	795	3.856	0.002

1.3.2 Anxiety estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	11.4	2.32	6.83	15.9
Angry	36.4	2.32	31.85	41.0
Bitter	48.2	2.32	43.68	52.8
Happy	15.3	2.32	10.74	19.8
Sad	50.5	2.32	45.94	55.0
Bittersweet	39.2	2.32	34.60	43.7

1.4.1 Sadness post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-30.506	2.373	795.000	-12.854	< .001
	-	Bitter	-64.719	2.373	795.000	-27.270	< .001
	-	Happy	2.144	2.373	795.000	0.903	0.946
	-	Sad	-78.406	2.373	795.000	-33.037	< .001
	-	Bittersweet	-48.363	2.373	795.000	-20.378	< .001
Angry	-	Bitter	-34.213	2.373	795.000	-14.416	< .001
	-	Happy	32.650	2.373	795.000	13.757	< .001
	-	Sad	-47.900	2.373	795.000	-20.183	< .001
	-	Bittersweet	-17.856	2.373	795.000	-7.524	< .001
Bitter	-	Happy	66.863	2.373	795.000	28.173	< .001
	-	Sad	-13.687	2.373	795.000	-5.767	< .001
	-	Bittersweet	16.356	2.373	795.000	6.892	< .001
Happy	-	Sad	-80.550	2.373	795.000	-33.940	< .001
	-	Bittersweet	-50.506	2.373	795.000	-21.281	< .001
Sad	-	Bittersweet	30.044	2.373	795.000	12.659	< .001

1.4.2 Sadness estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	7.900	1.801	4.366	11.434
Angry	38.406	1.801	34.872	41.941
Bitter	72.619	1.801	69.084	76.153
Happy	5.756	1.801	2.222	9.291
Sad	86.306	1.801	82.772	89.841
Bittersweet	56.263	1.801	52.728	59.797

1.5.1 Hope post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	54.96	2.41	795	22.791	< .001
	-	Bitter	53.61	2.41	795	22.229	< .001
	-	Happy	-10.66	2.41	795	-4.421	< .001
	-	Sad	52.41	2.41	795	21.731	< .001
	-	Bittersweet	10.98	2.41	795	4.554	< .001
Angry	-	Bitter	-1.36	2.41	795	-0.562	0.993
	-	Happy	-65.62	2.41	795	-27.212	< .001
	-	Sad	-2.56	2.41	795	-1.060	0.897
	-	Bittersweet	-43.98	2.41	795	-18.237	< .001
Bitter	-	Happy	-64.27	2.41	795	-26.650	< .001
	-	Sad	-1.20	2.41	795	-0.498	0.996
	-	Bittersweet	-42.63	2.41	795	-17.675	< .001
Happy	-	Sad	63.07	2.41	795	26.152	< .001
	-	Bittersweet	21.64	2.41	795	8.975	< .001
Sad	-	Bittersweet	-41.43	2.41	795	-17.177	< .001

1.5.2 Hope estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	63.06	1.83	59.46	66.7
Angry	8.09	1.83	4.50	11.7
Bitter	9.45	1.83	5.86	13.0
Happy	73.72	1.83	70.12	77.3
Sad	10.65	1.83	7.06	14.2
Bittersweet	52.08	1.83	48.48	55.7

1.6.1 Challenge post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	12.00	3.02	795	3.978	0.001
	-	Bitter	17.98	3.02	795	5.961	< .001
	-	Happy	-9.67	3.02	795	-3.208	0.017
	-	Sad	22.58	3.02	795	7.486	< .001
	-	Bittersweet	2.26	3.02	795	0.750	0.975
Angry	-	Bitter	5.98	3.02	795	1.983	0.353
	-	Happy	-21.67	3.02	795	-7.186	< .001
	-	Sad	10.58	3.02	795	3.508	0.006
	-	Bittersweet	-9.74	3.02	795	-3.228	0.016
Bitter	-	Happy	-27.66	3.02	795	-9.169	< .001
	-	Sad	4.60	3.02	795	1.525	0.648
	-	Bittersweet	-15.72	3.02	795	-5.211	< .001
Happy	-	Sad	32.26	3.02	795	10.694	< .001
	-	Bittersweet	11.94	3.02	795	3.958	0.001
Sad	-	Bittersweet	-20.32	3.02	795	-6.736	< .001

1.6.2 Challenge estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	35.7	2.36	31.01	40.3
Angry	23.7	2.36	19.01	28.3
Bitter	17.7	2.36	13.03	22.3
Happy	45.3	2.36	40.69	50.0
Sad	13.1	2.36	8.43	17.7
Bittersweet	33.4	2.36	28.75	38.0

1.7.1 Happiness post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Anger	76.4000	1.99	795	38.3596	< .001
	-	Bitter	75.9688	1.99	795	38.1431	< .001
	-	Happy	-10.9562	1.99	795	-5.5010	< .001
	-	Sad	76.3188	1.99	795	38.3188	< .001
	-	Bittersweet	24.7063	1.99	795	12.4047	< .001
Anger	-	Bitter	-0.4312	1.99	795	-0.2165	1.000
	-	Happy	-87.3562	1.99	795	-43.8606	< .001
	-	Sad	-0.0812	1.99	795	-0.0408	1.000
	-	Bittersweet	-51.6937	1.99	795	-25.9549	< .001
Bitter	-	Happy	-86.9250	1.99	795	-43.6441	< .001
	-	Sad	0.3500	1.99	795	0.1757	1.000
	-	Bittersweet	-51.2625	1.99	795	-25.7383	< .001
Happy	-	Sad	87.2750	1.99	795	43.8198	< .001
	-	Bittersweet	35.6625	1.99	795	17.9058	< .001
Sad	-	Bittersweet	-51.6125	1.99	795	-25.9141	< .001

1.7.2 Happiness estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	80.53	1.43	77.73	83.32
Anger	4.13	1.43	1.33	6.92
Bitter	4.56	1.43	1.76	7.35
Happy	91.48	1.43	88.68	94.28
Sad	4.21	1.43	1.41	7.00
Bittersweet	55.82	1.43	53.02	58.62

1.8.1 Pride post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	35.67	2.77	795	12.863	< .001
	-	Bitter	37.01	2.77	795	13.343	< .001
	-	Happy	-18.99	2.77	795	-6.849	< .001
	-	Sad	38.34	2.77	795	13.826	< .001
	-	Bittersweet	9.57	2.77	795	3.450	0.008
Angry	-	Bitter	1.33	2.77	795	0.480	0.997
	-	Happy	-54.67	2.77	795	-19.712	< .001
	-	Sad	2.67	2.77	795	0.962	0.930
	-	Bittersweet	-26.11	2.77	795	-9.413	< .001
Bitter	-	Happy	-56.00	2.77	795	-20.192	< .001
	-	Sad	1.34	2.77	795	0.482	0.997
	-	Bittersweet	-27.44	2.77	795	-9.893	< .001
Happy	-	Sad	57.34	2.77	795	20.674	< .001
	-	Bittersweet	28.56	2.77	795	10.299	< .001
Sad	-	Bittersweet	-28.78	2.77	795	-10.375	< .001

1.8.2 Pride estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	42.49	2.08	38.3975	46.58
Angry	6.81	2.08	2.7225	10.90
Bitter	5.48	2.08	1.3912	9.57
Happy	61.48	2.08	57.3912	65.57
Sad	4.14	2.08	0.0537	8.23
Bittersweet	32.92	2.08	28.8287	37.01

1.9.1 Gratitude post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	78.37	2.53	795	30.994	< .001
	-	Bitter	75.57	2.53	795	29.886	< .001
	-	Happy	-2.26	2.53	795	-0.892	0.948
	-	Sad	62.21	2.53	795	24.604	< .001
	-	Bittersweet	22.73	2.53	795	8.990	< .001
Angry	-	Bitter	-2.80	2.53	795	-1.107	0.878
	-	Happy	-80.62	2.53	795	-31.886	< .001
	-	Sad	-16.16	2.53	795	-6.390	< .001
	-	Bittersweet	-55.64	2.53	795	-22.004	< .001
Bitter	-	Happy	-77.83	2.53	795	-30.779	< .001
	-	Sad	-13.36	2.53	795	-5.282	< .001
	-	Bittersweet	-52.84	2.53	795	-20.897	< .001
Happy	-	Sad	64.47	2.53	795	25.497	< .001
	-	Bittersweet	24.99	2.53	795	9.882	< .001
Sad	-	Bittersweet	-39.48	2.53	795	-15.614	< .001

1.9.2 Gratitude estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	85.75	1.89	82.04	89.5
Angry	7.38	1.89	3.67	11.1
Bitter	10.18	1.89	6.47	13.9
Happy	88.01	1.89	84.30	91.7
Sad	23.54	1.89	19.83	27.2
Bittersweet	63.02	1.89	59.31	66.7

Supplementary Table 2. Post hoc comparisons for self-reported appraisals in study 2

2.1.1 Relevance post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	2.41	2.72	795	0.888	0.949
	-	Bitter	-12.47	2.72	795	-4.591	< .001
	-	Happy	-18.78	2.72	795	-6.916	< .001
	-	Sad	-17.64	2.72	795	-6.494	< .001
	-	Bittersweet	-14.45	2.72	795	-5.321	< .001
Angry	-	Bitter	-14.88	2.72	795	-5.480	< .001
	-	Happy	-21.19	2.72	795	-7.804	< .001
	-	Sad	-20.05	2.72	795	-7.383	< .001
	-	Bittersweet	-16.86	2.72	795	-6.209	< .001
Bitter	-	Happy	-6.31	2.72	795	-2.324	0.185
	-	Sad	-5.17	2.72	795	-1.903	0.401
	-	Bittersweet	-1.98	2.72	795	-0.730	0.978
Happy	-	Sad	1.14	2.72	795	0.421	0.998
	-	Bittersweet	4.33	2.72	795	1.595	0.602
Sad	-	Bittersweet	3.19	2.72	795	1.174	0.849

2.1.2 Relevance estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	61.4	2.04	57.4	65.5
Angry	59.0	2.04	55.0	63.0
Bitter	73.9	2.04	69.9	77.9
Happy	80.2	2.04	76.2	84.2
Sad	79.1	2.04	75.1	83.1
Bittersweet	75.9	2.04	71.9	79.9

2.2.1 Congruence post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	75.356	2.26	795	33.2956	< .001
	-	Bitter	75.513	2.26	795	33.3646	< .001
	-	Happy	-6.612	2.26	795	-2.9217	0.042
	-	Sad	72.513	2.26	795	32.0391	< .001
	-	Bittersweet	29.656	2.26	795	13.1034	< .001
Angry	-	Bitter	0.156	2.26	795	0.0690	1.000
	-	Happy	-81.969	2.26	795	-36.2173	< .001
	-	Sad	-2.844	2.26	795	-1.2565	0.808
	-	Bittersweet	-45.700	2.26	795	-20.1922	< .001
Bitter	-	Happy	-82.125	2.26	795	-36.2863	< .001
	-	Sad	-3.000	2.26	795	-1.3255	0.771
	-	Bittersweet	-45.856	2.26	795	-20.2612	< .001
Happy	-	Sad	79.125	2.26	795	34.9608	< .001
	-	Bittersweet	36.269	2.26	795	16.0251	< .001
Sad	-	Bittersweet	-42.856	2.26	795	-18.9357	< .001

2.2.2 Congruence estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	85.23	1.62	82.06	88.4
Angry	9.88	1.62	6.70	13.0
Bitter	9.72	1.62	6.55	12.9
Happy	91.84	1.62	88.67	95.0
Sad	12.72	1.62	9.55	15.9
Bittersweet	55.58	1.62	52.40	58.7

2.3.1 Incongruence post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-72.30	2.44	795	-29.67	< .001
	-	Bitter	-75.08	2.44	795	-30.81	< .001
	-	Happy	5.00	2.44	795	2.05	0.314
	-	Sad	-67.99	2.44	795	-27.90	< .001
	-	Bittersweet	-30.63	2.44	795	-12.57	< .001
Angry	-	Bitter	-2.78	2.44	795	-1.14	0.865
	-	Happy	77.30	2.44	795	31.72	< .001
	-	Sad	4.31	2.44	795	1.77	0.488
	-	Bittersweet	41.67	2.44	795	17.10	< .001
Bitter	-	Happy	80.08	2.44	795	32.86	< .001
	-	Sad	7.08	2.44	795	2.91	0.044
	-	Bittersweet	44.45	2.44	795	18.24	< .001
Happy	-	Sad	-72.99	2.44	795	-29.95	< .001
	-	Bittersweet	-35.62	2.44	795	-14.62	< .001
Sad	-	Bittersweet	37.37	2.44	795	15.33	< .001

2.3.2 Incongruence estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	14.28	1.71	10.92	17.6
Angry	86.58	1.71	83.22	89.9
Bitter	89.36	1.71	86.00	92.7
Happy	9.28	1.71	5.92	12.6
Sad	82.28	1.71	78.92	85.6
Bittersweet	44.91	1.71	41.55	48.3

2.4.1 Self-accountability post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	33.663	3.38	795	9.955	< .001
	-	Bitter	28.475	3.38	795	8.421	< .001
	-	Happy	4.700	3.38	795	1.390	0.733
	-	Sad	27.644	3.38	795	8.175	< .001
	-	Bittersweet	20.350	3.38	795	6.018	< .001
Angry	-	Bitter	-5.187	3.38	795	-1.534	0.642
	-	Happy	-28.962	3.38	795	-8.565	< .001
	-	Sad	-6.019	3.38	795	-1.780	0.479
	-	Bittersweet	-13.312	3.38	795	-3.937	0.001
Bitter	-	Happy	-23.775	3.38	795	-7.031	< .001
	-	Sad	-0.831	3.38	795	-0.246	1.000
	-	Bittersweet	-8.125	3.38	795	-2.403	0.156
Happy	-	Sad	22.944	3.38	795	6.785	< .001
	-	Bittersweet	15.650	3.38	795	4.628	< .001
Sad	-	Bittersweet	-7.294	3.38	795	-2.157	0.259

2.4.2 Self-accountability estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	60.5	2.51	55.6	65.4
Angry	26.8	2.51	21.9	31.8
Bitter	32.0	2.51	27.1	36.9
Happy	55.8	2.51	50.9	60.7
Sad	32.9	2.51	27.9	37.8
Bittersweet	40.1	2.51	35.2	45.1

2.5.1 Other-accountability post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-39.26	3.43	795	-11.452	< .001
	-	Bitter	-24.98	3.43	795	-7.287	< .001
	-	Happy	-11.99	3.43	795	-3.497	0.007
	-	Sad	11.25	3.43	795	3.281	0.014
	-	Bittersweet	2.88	3.43	795	0.839	0.960
Angry	-	Bitter	14.28	3.43	795	4.166	< .001
	-	Happy	27.28	3.43	795	7.956	< .001
	-	Sad	50.51	3.43	795	14.734	< .001
	-	Bittersweet	42.14	3.43	795	12.291	< .001
Bitter	-	Happy	12.99	3.43	795	3.790	0.002
	-	Sad	36.23	3.43	795	10.568	< .001
	-	Bittersweet	27.86	3.43	795	8.125	< .001
Happy	-	Sad	23.24	3.43	795	6.778	< .001
	-	Bittersweet	14.86	3.43	795	4.335	< .001
Sad	-	Bittersweet	-8.38	3.43	795	-2.443	0.143

2.5.2 Other-accountability estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	40.4	2.53	35.4	45.3
Angry	79.6	2.53	74.7	84.6
Bitter	65.4	2.53	60.4	70.3
Happy	52.4	2.53	47.4	57.3
Sad	29.1	2.53	24.2	34.1
Bittersweet	37.5	2.53	32.5	42.5

2.6.1 Future expectancy post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	52.644	3.02	795	17.418	< .001
	-	Bitter	50.419	3.02	795	16.682	< .001
	-	Happy	-6.375	3.02	795	-2.109	0.283
	-	Sad	51.225	3.02	795	16.948	< .001
	-	Bittersweet	23.081	3.02	795	7.637	< .001
Angry	-	Bitter	-2.225	3.02	795	-0.736	0.977
	-	Happy	-59.019	3.02	795	-19.527	< .001
	-	Sad	-1.419	3.02	795	-0.469	0.997
	-	Bittersweet	-29.562	3.02	795	-9.781	< .001
Bitter	-	Happy	-56.794	3.02	795	-18.791	< .001
	-	Sad	0.806	3.02	795	0.267	1.000
	-	Bittersweet	-27.338	3.02	795	-9.045	< .001
Happy	-	Sad	57.600	3.02	795	19.058	< .001
	-	Bittersweet	29.456	3.02	795	9.746	< .001
Sad	-	Bittersweet	-28.144	3.02	795	-9.312	< .001

2.6.2 Future expectancy estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	77.2	2.19	72.9	81.5
Angry	24.6	2.19	20.3	28.9
Bitter	26.8	2.19	22.5	31.1
Happy	83.6	2.19	79.3	87.9
Sad	26.0	2.19	21.7	30.3
Bittersweet	54.1	2.19	49.8	58.4

2.7.1 PFCP post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	27.03	3.10	795	8.719	< .001
	-	Bitter	34.31	3.10	795	11.065	< .001
	-	Happy	2.78	3.10	795	0.897	0.947
	-	Sad	35.76	3.10	795	11.533	< .001
	-	Bittersweet	20.76	3.10	795	6.697	< .001
Angry	-	Bitter	7.28	3.10	795	2.346	0.177
	-	Happy	-24.25	3.10	795	-7.822	< .001
	-	Sad	8.73	3.10	795	2.814	0.056
	-	Bittersweet	-6.27	3.10	795	-2.022	0.331
Bitter	-	Happy	-31.53	3.10	795	-10.168	< .001
	-	Sad	1.45	3.10	795	0.468	0.997
	-	Bittersweet	-13.54	3.10	795	-4.368	< .001
Happy	-	Sad	32.98	3.10	795	10.636	< .001
	-	Bittersweet	17.98	3.10	795	5.800	< .001
Sad	-	Bittersweet	-14.99	3.10	795	-4.836	< .001

2.7.2 PFCP estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	59.4	2.30	54.9	63.9
Angry	32.4	2.30	27.9	36.9
Bitter	25.1	2.30	20.6	29.6
Happy	56.6	2.30	52.1	61.1
Sad	23.6	2.30	19.1	28.1
Bittersweet	38.6	2.30	34.1	43.1

2.8.1 AFCP post hoc comparisons

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	22.500	2.63	795	8.547	< .001
	-	Bitter	28.169	2.63	795	10.701	< .001
	-	Happy	2.000	2.63	795	0.760	0.974
	-	Sad	28.525	2.63	795	10.836	< .001
	-	Bittersweet	16.863	2.63	795	6.406	< .001
Angry	-	Bitter	5.669	2.63	795	2.153	0.261
	-	Happy	-20.500	2.63	795	-7.788	< .001
	-	Sad	6.025	2.63	795	2.289	0.200
	-	Bittersweet	-5.637	2.63	795	-2.142	0.267
Bitter	-	Happy	-26.169	2.63	795	-9.941	< .001
	-	Sad	0.356	2.63	795	0.135	1.000
	-	Bittersweet	-11.306	2.63	795	-4.295	< .001
Happy	-	Sad	26.525	2.63	795	10.076	< .001
	-	Bittersweet	14.863	2.63	795	5.646	< .001
Sad	-	Bittersweet	-11.663	2.63	795	-4.430	< .001

2.8.2 AFCP estimated marginal means

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	87.8	2.13	83.6	92.0
Angry	65.3	2.13	61.1	69.5
Bitter	59.6	2.13	55.4	63.8
Happy	85.8	2.13	81.6	90.0
Sad	59.3	2.13	55.1	63.4
Bittersweet	70.9	2.13	66.7	75.1

Supplementary Table 3. Estimated marginal means for voice-based emotion detection

3.1 Predicted calmness

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	0.493	0.0111	0.471	0.514
Angry	0.458	0.0111	0.437	0.480
Bitter	0.515	0.0111	0.493	0.537
Happy	0.464	0.0111	0.442	0.486
Sad	0.540	0.0111	0.518	0.562
Bittersweet	0.520	0.0111	0.498	0.542

3.2 Predicted anger

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	0.504	0.0130	0.479	0.530
Angry	0.526	0.0130	0.500	0.552
Bitter	0.499	0.0130	0.474	0.525
Happy	0.514	0.0130	0.488	0.539
Sad	0.470	0.0130	0.445	0.496
Bittersweet	0.476	0.0130	0.451	0.502

3.3 Predicted happiness

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	0.450	0.0104	0.430	0.471
Angry	0.534	0.0104	0.514	0.554
Bitter	0.507	0.0104	0.487	0.528
Happy	0.527	0.0104	0.507	0.547
Sad	0.468	0.0104	0.447	0.488
Bittersweet	0.510	0.0104	0.490	0.530

3.4 Predicted sadness

			95% Confidence Interval	
Story	Mean	SE	Lower	Upper
Calm	0.511	0.0130	0.486	0.537
Angry	0.462	0.0130	0.436	0.487
Bitter	0.490	0.0130	0.464	0.515
Happy	0.470	0.0130	0.444	0.495
Sad	0.548	0.0130	0.523	0.574
Bittersweet	0.522	0.0130	0.497	0.548

Supplementary Table 4. Post hoc comparisons for voice-based emotion detection among singular emotion stories

4.1 Predicted calmness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	0.034	0.019	156.000	1.779	0.287
	-	Happy	0.029	0.016	156.000	1.743	0.305
	-	Sad	-0.047	0.017	156.000	-2.845	0.026
Angry	-	Happy	-0.006	0.019	156.000	-0.301	0.990
	-	Sad	-0.081	0.017	156.000	-4.714	< .001
Happy	-	Sad	-0.076	0.017	156.000	-4.589	< .001

4.2 Predicted anger

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-0.022	0.020	156.000	-1.066	0.711
	-	Happy	-0.010	0.021	156.000	-0.465	0.967
	-	Sad	0.034	0.023	156.000	1.506	0.436
Angry	-	Happy	0.012	0.020	156.000	0.619	0.926
	-	Sad	0.056	0.022	156.000	2.544	0.057
Happy	-	Sad	0.044	0.020	156.000	2.221	0.122

4.3 Predicted happiness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	-0.084	0.017	156.000	-5.031	< .001
	-	Happy	-0.077	0.016	156.000	-4.824	< .001
	-	Sad	-0.017	0.017	156.000	-1.036	0.728
Angry	-	Happy	0.007	0.016	156.000	0.446	0.970
	-	Sad	0.066	0.015	156.000	4.282	< .001
Happy	-	Sad	0.059	0.016	156.000	3.748	0.001

4.4 Predicted sadness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Calm	-	Angry	0.050	0.020	156.000	2.440	0.074
	-	Happy	0.041	0.021	156.000	2.011	0.188
	-	Sad	-0.037	0.022	156.000	-1.669	0.344
Angry	-	Happy	-0.008	0.020	156.000	-0.423	0.974
	-	Sad	-0.087	0.022	156.000	-3.887	< .001
Happy	-	Sad	-0.078	0.020	156.000	-4.000	< .001

Supplementary Table 5. Post hoc comparisons for voice-based emotion detection between blend stories and blend constituting singular emotion stories

5.1 Bitter – angry – sad on predicted anger

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Bitter	-	Angry	-0.027	0.019	156.000	-1.406	0.340
	-	Sad	0.029	0.018	156.000	1.598	0.249
Angry	-	Sad	0.056	0.022	156.000	2.544	0.032

5.2 Bitter – angry – sad on predicted sadness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Bitter	-	Angry	0.028	0.019	156.000	1.470	0.308
	-	Sad	-0.059	0.018	156.000	-3.322	0.003
Angry	-	Sad	-0.087	0.022	156.000	-3.887	< .001

5.3 Bittersweet – sad – happy on predicted sadness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Bittersweet	-	Sad	-0.026	0.021	156.000	-1.255	0.423
	-	Happy	0.052	0.019	156.000	2.772	0.017
Sad	-	Happy	0.078	0.020	156.000	4.000	< .001

5.4 Bittersweet – sad – happy on predicted happiness

Comparison							
Story		Story	Mean Difference	SE	df	t	p_{tukey}
Bittersweet	-	Sad	0.042	0.016	156.000	2.658	0.023
	-	Happy	-0.017	0.015	156.000	-1.113	0.507
Sad	-	Happy	-0.059	0.016	156.000	-3.748	< .001

REFERENCES

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755-1762.

Atias, D., & Aviezer, H. (2020). Real-life and posed vocalizations to lottery wins differ fundamentally in their perceived valence. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0000931

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in Psychology*, *6*, 428.

Boersma, P. & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 6 June 2020 from http://www.praat.org/.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin*, *115*(3), 401-423.

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., Brooks, K. P., & Nesselroade, J. R. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging, 26*(1), 21–33.

Charles, S. T., Piazza, J. R., & Urban, E. J. (2017). Mixed emotions across adulthood: when, where, and why?. *Current opinion in behavioral sciences*, *15*, 58-61.

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

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & emotion*, *27*(4), 658-684.

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*, *74*(6), 698.

Croes, E. A. J., Antheunis, M. L., Schouten, A. P., & Krahmer, E. J. (2019). Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of Social and Personal Relationships*, *36*(4), 1210–1232.

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological science*, *7*(3), 181-185.

Ellsworth, P. C., & Smith, C. A. (1988a). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and emotion*, *12*(3), 271-302.

Ellsworth, P. C., & Smith, C. A. (1988b). Shades of joy: Patterns of appraisal differentiating pleasant emotions. *Cognition & Emotion*, *2*(4), 301-331.

Ellsworth, P. C., & Tong, E. M. (2006). What does it mean to be angry at yourself? Categories, appraisals, and the problem of language. *Emotion*, *6*(4), 572-586.

Ersner-Hershfield, H., Mikels, J. A., Sullivan, S. J., & Carstensen, L. L. (2008). Poignancy: Mixed emotional experience in the face of meaningful endings. *Journal of Personality and Social Psychology*, *94*(1), 158-167.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, *56*(3), 218-226.

Friedrich, S., Konietschke, F., & Pauly, M. (2021). MANOVA.RM: Resampling-Based Analysis of Multivariate Data and Repeated Measures Designs. R package version 0.4.3. https://CRAN.R-project.org/package=MANOVA.RM

Gross, J. J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, *3*(1), 8-16.

Grossmann, I., Huynh, A. C., & Ellsworth, P. C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *Journal of Personality and Social Psychology, 111*(6), 895–916.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, *22*(2), 327-352.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, *99*(3), 561-565.

The jamovi project (2021). *jamovi* (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381-412.

Kuppens, P., Van Mechelen, I., Smits, D. J., & De Boeck, P. (2003). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion*, *3*(3), 254-269.

Kuppens, P., Van Mechelen, I., Smits, D. J., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition and Emotion*, *21*(4), 689-713.

Labouvie-Vief, G., Lumley, M. A., Jain, E., & Heinze, H. (2003). Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion*, *3*(2), 115-126.

Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology*, *100*(6), 1095-1110.

Larsen, J. T., & McGraw, A. P. (2014). The case for mixed emotions. *Social and Personality Psychology Compass*, *8*(6), 263-274.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time?. *Journal of Personality and Social Psychology*, *81*(4), 684-696.

Ledoux, J. (1991). Emotion and the brain. *Journal of NIH research*, *3*, 49-51.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.

Man, V., Nohlen, H. U., Melo, H., & Cunningham, W. A. (2017). Hierarchical brain systems support multiple representations of valence and mixed affect. *Emotion Review, 9*(2), 124-132.

Oatley, K., & Johnson-Laird, P. (1996). The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction. In L. L. Martin, & A. Tesser (Eds.), *Striving and feeling: Interactions among goals, affect, and self-regulation* (pp. 363-393, Chapter x, 408 Pages). Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Öhman, A. (1999). Distinguishing unconscious from conscious emotional processes: Methodological considerations and theoretical implications. In T. Dalgleish, & M. J.

Power (Eds.), *Handbook of cognition and emotion* (pp. 321-352, Chapter xxi, 843 Pages). John Wiley & Sons Ltd, New York, NY.

Rao, K. S., Koolagudi, S. G., & Vempada, R. R. (2013). Emotion recognition from speech using global and local prosodic features. *International journal of speech technology*, *16*(2), 143-160.

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin*, *27*(11), 1520-1532.

Roseman, I. J., Spindel, M. S., & Jose, P. E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology*, *59* (5), 899-915.

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2019). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978.

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

Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review, 9*(2), 111-117.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological bulletin*, *125*(1), 3-30.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and emotion*, *15*(2), 123-148.

Schimmack, U., & Colcombe, S. (2007). Eliciting mixed feelings with the paired-picture paradigm: A tribute to Kellogg (1915). *Cognition and Emotion*, *21*(7), 1546-1553.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*, *8*(1), 289-317.

Smith, C. A. (1991). The self, appraisal, and coping. In C. R. Snyder, & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective* (pp. 116-137). Pergamon Press, Elmsford, NY.

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

Smith, C. A., & Ellsworth, P. C. (1987). Patterns of appraisal and emotion related to taking an exam. *Journal of Personality and Social Psychology*, *52*(3), 475-488.

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In L. A. Pervin (Ed.), *Handbook of personality theory and research* (pp. 609-637). New York: Guilford.

Smith, C. A., Tong, E. M. W., & Ellsworth, P. C. (2014). The differentiation of positive emotional experience as viewed through the lens of appraisal theory. In M. M. Tugade, M. N. Shiota & L. D. Kirby (Eds.), *Handbook of positive emotions* (pp. 11-27, Chapter xv, 527 Pages). Guilford Press, New York, NY.

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of psycholinguistic research*, *28*(4), 347-365.

Teigen, K. H. (1997). Luck, envy and gratitude: It could have been different. *Scandinavian Journal of Psychology*, *38*(4), 313-323.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality*, 34(2), 178-197.