Revision submission date: 04-03-2018

Interaction and threshold effects of appraisal on componential patterns of emotion: A study using cross-cultural semantic data

Ben Meuleman, Agnes Moors, Johnny Fontaine, Olivier Renaud & Klaus R. Scherer

Author affiliations

Ben Meuleman, Swiss Center for Affective Sciences, University of Geneva; Agnes Moors, Department of Quantitative Psychology and Individual Differences an Centre for Social and Cultural Psychology, Faculty of Psychology and Educational Sciences, KU Leuven; Johnny Fontaine, Department of Personnel Management, Work and Organizational psychology, Universiteit Gent; Olivier Renaud, Department of Methodology and Data Analysis, Section of Psychology, University of Geneva; Klaus Scherer, Swiss Center for Affective Sciences, University of Geneva

Correspondence

Ben Meuleman

Swiss Center for Affective Sciences (CISA) – University of Geneva

Telephone: +41 (0)22 379 09 79

ben.meuleman@unige.ch

Campus Biotech

CISA – University of Geneva

Chemin des Mines 9

CH-1202 Genève

Abstract

Studies that investigated the relation between appraisal and emotion have largely focused on the linear effect of appraisal criteria on subjective feelings (e.g., the effect of appraised goal obstruction on anger). Emotional responding can be extended to include more than just feelings, however. Componential definitions of emotion also add motivation, physiology, and expression. Moreover, a linear model is not compatible with the idea held by many appraisal theorists that appraisal criteria *interact* to produce emotional responding. In the present study, we modelled adaptive non-linear interaction effects of appraisal criteria on motivation, expression, and physiology simultaneously. We applied a combination of principal component analysis (PCA) for data reduction, and multivariate adaptive regression splines (MARS) for automatic interaction identification. Data were obtained from a large-scale cross-cultural study on emotion concepts conducted in 27 countries, which represented semantic profiles of component information in 24 common emotion words. Results of modelling indicated that **(a)** appraisal of relevance, familiarity, goal compatibility, coping potential, and suddenness showed main effects on component responses, **(b)** appraisals of agency and norm compatibility uniquely showed interaction effects on component responses, **(c)** interaction effects explained significant variance only in some component responses but not all, and **(d)** the emotion patterns simulated by the fitted MARS model could be clustered according to qualitative emotion categories.

Keywords: emotion; appraisal; machine learning; cross-cultural psychology; computational modelling

1 Introduction

In the past decades, emotion theories in psychology have converged on what is known as a componential view of emotion. According to this view, an emotion is a temporary episode of coordinated components, that is, changes in five organismic subsystems (Frijda, 2008; Frijda & Scherer, 2009; Moors, 2009; Scherer, 2005): (a) an appraisal component that involves changes in the evaluation of an event or stimulus with respect to its personal significance, (b) a motivational component that involves changes in action tendencies such as the tendencies to attack and avoid, (c) a physiological component that involves changes in autonomic physiological responses of the body such as sweating, increases in heart rate, or changes in blood pressure, (d) an expression component that involves changes in expressive behaviour in the face (e.g., smiling), the voice (e.g., screaming), or the body (e.g., attacking), and (e) a feeling component that involves the conscious experience of all the other components integrated in a kind of Gestalt. This experience can (but does not have to) be categorized or verbally labelled (e.g., as feelings of anger, fear, or sadness; see Figure 1 for a schematic depiction of all components).

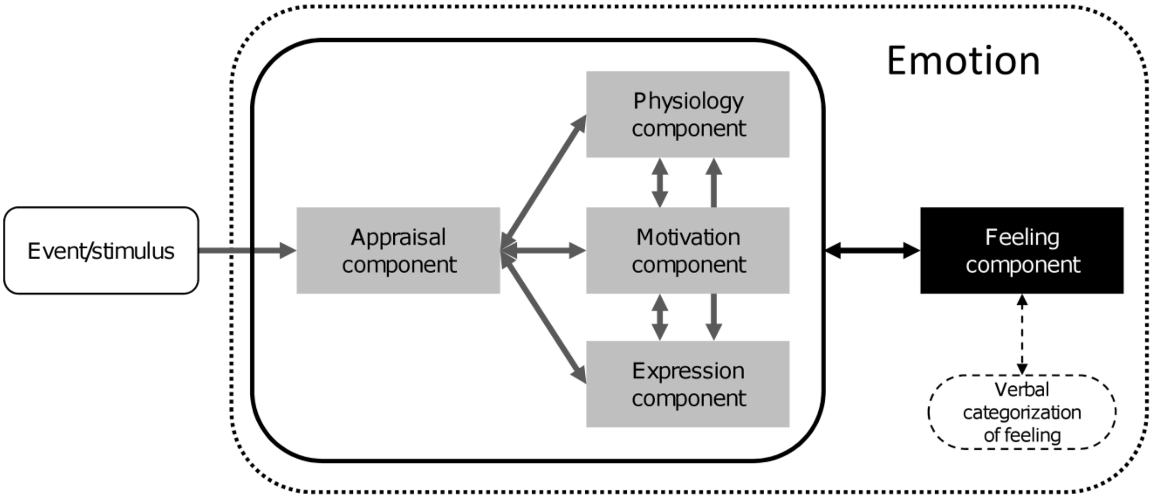

Figure 1. Emotion as a coordinated set of components: appraisal, motivation, physiology, expression, and feeling. Appraisal theories posit that appraisal of an event causes changes in motivation, physiology, and expression. Jointly, the changes in these four organismic subsystems enter conscious experience as a feeling, which may be categorized with verbal emotion labels.

The componential view of emotion lists the components that are part of an emotional episode and introduces assumptions about the way in which changes in organismic subsystems are elicited and differentiated and about the relations between the components. Contemporary appraisal theories of emotion (Scherer, Schorr & Johnstone, 2001) are one type of componential theory. They place special emphasis on the appraisal component to explain the emotion process arguing that it is the appraisal of the event rather than the intrinsic properties of the event that drives the changes in the other four components. In Figure 1, this assumption has been represented by the absence of arrows leading directly from the emotion eliciting event to the motivation, physiology, expression, and feeling components. All changes in these components are causally mediated by the appraisal process (Moors, 2009; 2013).

Appraisal theories put forward a limited set of criteria, such as goal relevance, compatibility with goals and desires, compatibility with expectations, causal agency, potential to cope with or control the situation, and compatibility with norms and values (see Ellsworth & Scherer, 2003, for lists of common criteria). Combinations of evaluations on these criteria are thought to differentiate patterns of emotional responding in the other components. Some but not all of these patterns correspond well with colloquial emotion categories such as fear, anger, and joy. The idea that a handful of criteria can account for much of the variation in emotional responding has had obvious appeal to scientific study. It has stimulated a wealth of observational, experimental, and simulation research (see Ellsworth & Scherer, 2003, Moors & Scherer, 2013, and Marsella, Gratch, & Petta, 2010, for partial overviews of these three domains). All these studies support the existence of relations between appraisal criteria and emotional responding. At the same time, other important assumptions of appraisal theories have not been consistently studied in empirical research. In the present study, we wish to focus on two of these assumptions: *(a)* that emotional responding consists of changes in motivation, physiology, and expression, in addition to feeling—as per the componential definition—and *(b)* that appraisal criteria *interact* to differentiate emotional responding. In the next two sections, we take a closer look at these two assumptions and review how both theory and research have dealt with them. After that, we show how multicomponentiality and interactions can be studied using appropriate data and flexible statistical models.

1.1 Multicomponential responses

Appraisal theories agree that emotions are multicomponential phenomena and that the appraisal process should affect changes in motivation, physiology, expression, *and* feeling (see Figure 1). Nevertheless, empirical work on appraisal theory has overwhelmingly focused on explaining changes in feeling—usually discrete categories thereof. Such an approach neglects the motivation, physiology, and expression components, and reduces feelings to categorical labels. An empirical model of emotion that includes full componential information is desirable, firstly, for its comprehensiveness—allowing emotions to represented in much richer detail than in one- or two-component models—and, secondly, to investigate in a systematic fashion the link between different components of emotion. In this section we discuss how appraisal theories have dealt with multi-componentiality in emotion, both in theory and research.

The feeling component has a special status in the componential perspective on emotion, in that it is considered to be an integrated awareness of the changes in the other four components (Moors, 2009; Moors & Scherer, 2013; Scherer, 2005). Just as points, lines, and curves are combined into a global percept or Gestalt rather than as separate elements, changes in appraisal, motivation, physiology, and expression might combine to produce an integrated feeling that could receive a label. The number of possible feeling states is thought to be infinite. However, some patterns of changes and their associated feelings may occur more frequently than others (i.e., so-called modal emotions), leading to special words in our vocabulary to describe these patterns, such as joy, fear, and anger (Scherer, 2001, 2009).

One important class of appraisal theories primarily seeks to explain variation in labelled feeling categories on the basis of appraisal criteria (e.g., Lazarus, 1991; Roseman 2013). Another important class of appraisal theories seeks to explain variation in all components, without necessarily linking the patterns to feeling labels (e.g., Scherer, 2009). As examples of the latter, Frijda (1988) has proposed hypotheses about the influence of specific appraisals on specific action tendencies, while Scherer (2001; 2009) has proposed hypotheses about the influence of specific appraisals on specific physiological responses and facial action units. These two classes of theories have been identified by Moors (2014) as two *flavors* of appraisal theories.

When considering these two flavors of theories for empirical emotion modelling, it should be clear that the first flavor offers a less comprehensive characterization of an emotion episode than does the second. A model that only outputs feeling labels provides no explicit information on motivational, physiological, or expression changes. These changes are only *implied* by our own background knowledge of that label, for example, knowledge that the word "anger" conveys a possible tendency to attack, increased sweating, or lowered eyebrows. On the other hand, a model that outputs responses of all emotion components would not only be more comprehensive, it would also allow a more detailed decomposition of feelings. The diagram in Figure 1 implies that bodily and behavioral information co-determine changes in feeling. If we wish to find out which changes in which components emerge into consciousness, full component information would be required. For this reason, the appraisal theories of the second flavor promise to yield a more comprehensive understanding of emotions. Second-flavor appraisal research can be split into bi-componential and multi-componential research.

Bi-componential research investigates links between appraisal, on the one hand, and changes in motivation, physiology, or expression, on the other hand. Studies provide support for links between appraisals and (a) motivation changes, such as tendencies to attack or repair (e.g., Bossuyt, Moors, & De Houwer, 2014a,b; Fast & Chen, 2009; Frijda, Kuipers, & ter Schure, 1988; Moors & De Houwer, 2001; Nelissen & Zeelenberg, 2009), (b) physiological changes, such as cardiac, respiratory, temperature, and electrodermal responses (e.g., Aue, Flykt, & Scherer, 2007; Delplanque et al., 2009; Gentsch, Grandjean, & Scherer, 2013; Gentsch, Grandjean, & Scherer, 2014; Grandjean & Scherer, 2008; Kreibig, Gendolla, & Scherer, 2010; Kreibig, Gendolla, & Scherer, 2012; Lanctot & Hess, 2007; van Reekum et al., 2004), and (c) expression changes, such as facial muscle activity, vocal activity, and body posture (see reviews in chapters 15–17 in Scherer et al., 2001 and Scherer, 2009).

Multi-componential research examines hypotheses that involve three, four, or even all five components simultaneously. Appraisal theories (Frijda, 1993; Frijda, Kuipers, & ter Schure, 1989; Scherer, 2009) typically propose that effects of the appraisal component on the physiological and expression component are *mediate*d by the motivation component. In turn, all these components mediate between appraisal and the feeling component. Testing such hypotheses requires the modelling of at least three emotion components simultaneously, that is, a predictor, a mediator, and a response component. At present, such analyses remain scarce.

If there are indeed strong causal relations among components, they should show a high degree of synchronization. Scherer (2009) even proposes *component synchronization* as a defining criterion for demarcating emotional from non-emotional episodes (see also Lewis, 2005). He argues that strong synchronization or coherence among changes in appraisal, motivation, physiology and expression should increase the likelihood for an integrated feeling sensation to enter consciousness, for that feeling to be intense, and for the overall pattern of component changes to correspond to a modal emotion. Again, testing such hypotheses requires a multicomponential analysis.

To date, a handful of studies have collected data on all components of emotion (e.g., Fitness & Fletcher, 1993; Scherer & Wallbott, 1994) but few of these have analysed the data simultaneously in a common model. Nguwi and Cho (2012) re-analysed the ISEAR dataset (International Survey on Emotion and Antecedent Relations; Scherer & Wallbott, 1994) and integrated survey data on all five emotion components for seven common emotion categories into a single support vector classifier. Fontaine, Scherer, and Soriano (2013) conducted a cross-cultural survey on emotion knowledge similar to the ISEAR study, measuring all components of emotion across 144 individual questionnaire items for 24 common feeling categories (e.g., pride, guilt). Both of these studies achieved an overall discrimination performance of over 80% correct classification using information from all emotion components simultaneously. These results point to the relatively high coherence among different emotion components, and suggest the possibility of further and more complex analyses.

1.2 Interactions and nonlinear associations in appraisal theory

We have highlighted in the previous section that the five components of emotion are expected to change in a mutually coherent fashion during an emotion episode. In Figure 1 this has been represented by arrows that connect all components. So far, we have not addressed the exact nature of those arrows. Appraisal theorists have put forward complex relations between appraisal criteria and emotional responding. Nonlinear associations have been proposed in (at least) three different areas: interaction effects between appraisal criteria to differentiate emotional responding (Ortony, Collins & Clore, 1988; Lazarus, 2001; Roseman, 2001; Scherer, 2001), curvilinear associations between appraisal criteria and emotion intensity (Kappas, 2001; Tong, Ellsworth & Bishop, 2009), and time-dependent feedback processes (Lewis, 2005; Scherer, 2001, 2009). We limit our discussion to the first two kinds because the data in the present study do not allow us to analyse associations of the third kind.

Two appraisal criteria interact when the effect of one appraisal criterion on emotional responding is modified by levels of another appraisal criterion. Suppose that someone spills coffee on your friend at a restaurant. Depending on the agent of this spilling (you or the waiter), a different emotion may occur (e.g., guilt or anger). In this case, the appraisal of an undesirable outcome (coffee being spilled) is modified by the appraisal of agency (self vs. other agency) to differentiate the two emotions. All major appraisal theories predict this kind of interaction, including the theories of Ortony and colleagues (1988; the OCC model), Lazarus (1993; 2001), Roseman (2001; 2009), and Scherer (2001; 2009; the Componential Process Model). Interaction hypotheses in these theories usually take the form of configurational appraisal profiles associated with discrete emotion categories. Such configurations imply that an emotion is not just differentiated by a single appraisal criterion but by a joint combination of multiple appraisal criteria. This is most evident in the OCC model, which differentiates emotions according to a tree of decision rules (see 1988, p. 19). Paths in the tree correspond to configurations of appraisal values associated with specific emotions. Roseman (2001, pp. 70–71) and Scherer (2001, pp. 114–115) have proposed configurations in the form of contingency tables, with different emotion labels predicted at the intersection of different appraisal criteria values. These two types of representations—trees and tables—do not reflect a fundamentally different interaction structure, however. A contingency table could be rewritten as a decision tree, and vice-versa.Trees and tables highlight the depth of the interaction structure that is implied by appraisal theories. In Roseman's model (2001, pp. 70–71), for example, a classification of fear requires a six-way interaction of EXPECTEDNESS × AGENCY × GOAL COMPATIBILITY × COPING × CERTAINTY. In addition, decision trees make explicit the hierarchical nature of the interaction structure. Decision trees contain only one traditional main effect, which is the first decision rule at the top of the tree (e.g., if situation is appraised on non-relevant, output no emotion; if situation is appraised as relevant, move to next appraisal decision rule. All other effects in the tree are conditional on this main effect, and hence should be considered as pure modifiers (Hastie, Friedman & Tibshirani, 2007). This distinction has also been addressed by many appraisal theories. For example, most theories consider appraisal of (GOAL) RELEVANCE the main effect for determining whether an emotion is elicited at all, whereas appraisal of GOAL COMPATIBILITY is considered the main effect for separating positively from negatively valenced emotions (Ellsworth & Scherer, 2003). Other appraisal criteria are typically considered as modifiers of these main effects, such as appraisal of AGENCY, COPING POTENTIAL, and NORM COMPATIBILITY (Lazarus, 2001; Roseman, 2001; Scherer, 2001).

In addition to standard interaction effects between different appraisal criteria, some authors have proposed that an appraisal criterion can interact *with itself* to affect emotional responding (Kappas, 2001; Tong, Ellsworth, & Bishop, 2009). This happens when the relation between an appraisal criterion and emotional responding changes for different values of that appraisal criterion. This is true by definition for any nonlinear function (e.g., parabolas, exponential curves) where the rate of change is variable. Kappas (2001) argued that relations between appraisal criteria and emotion intensity should be expected to be *sigmoidal.* A sigmoidal function is an s-shaped curve that makes a smooth monotone transition between a lower and an upper output limit (see Figure 3 for an example). For appraisal theory, this corresponds to an assumption that emotion intensity should be bounded to certain limits, and should not increase indefinitely for increasing appraisal input. Sigmoidal curves can also be viewed as continuous approximations of binary threshold functions[1], and they share the property that output cannot increase substantially until input reaches a certain value.

Evidence for interaction effects between different appraisal criteria has been collected in a non-systematic fashion, encompassing both experimental as wel as observational studies. Experimental research has tested interactions in traditional factorial ANOVA designs, where levels of one or multiple appraisal criteria are manipulated across levels of other appraisal criteria in an orthogonal fashion. Two-way interaction effects of RELEVANCE × GOAL COMPATIBILITY have been reported on self-reported feelings of joy, sadness, and frustration (Kreibig, Gendolla, & Scherer, 2012; Nyer, 1997), on cardiovascular and respiratory measures of physiology (Kreibig, Gendolla, & Scherer, 2012), and on electromyogram corrugator measures of expression (Aue, Flykt, & Scherer, 2007). Trends toward two-way interaction effects for GOAL COMPATIBILITY × COPING were found in two

[1] Also known as the Heaviside step function

studies conducted by Gentsch, Grandjean and Scherer (2013; 2014) for measures of electroencephalogram P300 mean amplitude and electromyogram zygomaticus activity, respectively. A three-way interaction effect of RELEVANCE × GOAL COMPATIBILITY × COPING was reported by Nyer (1997) on self-reported feelings of anger. These findings lend some support to the interaction hypotheses of appraisal theories but in many experimental studies (including those just cited), the majority of interaction tests that were performed failed to reach significance (e.g., Kreibig, Gendolla, & Scherer, 2012; McGraw, 1987; Russell & McAuley, 1986; van Reekum et al., 2004). For example, van Reekum and colleagues (2004) manipulated appraisals of PLEASANTNESS and GOAL COMPATIBILITY in a videogame paradigm but found no two-way interaction across five self-reported feelings (pride, joy, anger, surprise, tenseness), four physiological measures (skin conductance, interbeat interval, pulse transit time, finger temperature slope), and two expression measures (frontalis and forearm extensor activity).

A potential drawback of traditional ANOVAs is that the number of conditions grows exponentially with the number of appraisal criteria. For example, for six appraisal criteria (an average across theories, see Ellsworth & Scherer, 2003) with a binary data format (0 = appraisal absent, 1 = appraisal present), a full ANOVA design requires 64 conditions. With a ternary format (-1 = appraisal negative, 0 = appraisal absent, 1 = appraisal positive), a full ANOVA design requires 729 conditions. Many of these conditions may be irrelevant or difficult to sample. This makes traditional ANOVAs not well-suited for testing the kind of sixth-order interactions predicted by Roseman (2001) or Scherer (2001) and may explain why they have not been systematically investigated.

Observational studies typically investigate the relation between a set of appraisal criteria and features of subjective feeling (e.g., fear intensity, anger intensity, joy intensity) using standard linear models, such as additive linear regression[2], logistic regression, and linear discriminant analysis. Studies that used this approach have been conducted by Smith and Ellsworth (1985), McGraw (1987), Frijda, Kuipers, and ter Schure (1989), Reisenzein and Spielhofer (1994), Sonnemans and Frijda

[2] Including Pearson correlation, *t*-test, ANCOVA, and non-factorial ANOVA

(1995), Scherer and Ceschi (1997), Scherer (1997), Ellsworth and Smith (1998), Ruth, Brunel, and Ötnes (2002), Kuppens et al. (2003), Tong et al. (2005), Siemer, Mauss, and Gross (2007), Hosany (2011), and Brans and Verduyn (2014). These analyses typically yield tables of one-to-one relations between appraisal criteria and feeling features (e.g., correlation coefficients, regression weights), as well as an index of explained variance or discriminative power. This approach has been useful for identifying specific appraisal determinants of specific emotions. Yet it is less suitable to measure the interaction hypotheses of appraisal theories.

Two advantages that observational studies have over experimental studies is that they can include more appraisal criteria than a factorial ANOVA would allow, and that they do not have to restrict the format of the criteria to binary or ternary format. The disadvantage of this approach is the difficulty of selecting meaningful interaction effects in standard linear models. For six appraisal criteria, there are 15 possible two-way interactions to construct, 20 possible three-way interactions, 15 possible four-way interactions, 6 possible five-way interactions, and 1 six-way interaction. Without automated procedures or prior hypotheses, testing all these effects simultaneously would not only strain statistical power but also raise issues of multiple testing. Nevertheless, a handful of observational studies, however, has attempted to model interactions. McGraw (1987) and Ruth et al. (2002) added two-way interaction terms to a standard linear model, although they could not significantly improve the resulting fit of the models. Graham, Kowalski and Crocker (2002) found a significant two-way interaction for RELEVANCE × GOAL COMPATIBILITY on self-reported sadness feelings. In addition, they found two three-way interactions, one for RELEVANCE × GOAL COMPATIBILITY × AGENCY and one for RELEVANCE × GOAL COMPATIBILITY × CONTROL on self-reported self-assurance feelings. However, the authors cast doubt on these findings due to the large number of exploratory tests. Tong et al. (2005) tested a configurational hypothesis of three appraisal criteria on self-reported happiness intensity, and found a significant three-way interaction of PLEASANTNESS × PERCEIVED CONTROL × MORAL VIOLATION on happiness ratings (using ecological momentary assessment). Importantly, PLEASANTNESS was the only criterion to show a main effect on happiness ratings in addition to its interaction effect. The criteria PERCEIVED CONTROL and MORAL VIOLATION, on the other hand, were pure modifiers. Such effects support the hierarchical interaction structure suggested by decision trees and prediction tables from appraisal theories. In a follow-up study, Tong and colleagues (2009) also tested sigmoidal relations between appraisal criteria and emotion intensity. Significant curvilinear associations were found between seven appraisal criteria and five emotions across different subject pools, cultures, and sampling methods.

Meuleman and Scherer (2013) attempted to address some of the disadvantages of standard linear models by conducting an automated search for interactions in a large observational dataset on recalled emotion episodes. They modelled a categorical emotion variable (12 levels reflecting feeling labels such as pride, joy, rage, and despair) as a function of 25 appraisal criteria. Using classification techniques from the field of machine learning, it was found that interactive models significantly outperformed additive (linear) models. Furthermore, the interaction structure was found to be moderately hierarchical, with a main effect of the GOAL COMPATIBILITY appraisal separating positively from negatively valenced emotions, and interaction effects involving the appraisals of AGENCY and NORM COMPATIBILITY separating emotions within those two clusters (e.g., anger vs. fear). Meuleman and Scherer (2013) advocated the use of machine learning to estimate associations between appraisal criteria and emotional responding without assumptions of linearity. We will also adopt this perspective in the present paper, and show how the use of a more complex model does not necessarily compromise interpretability.

1.3 Present Study

The aim of the present study was to analyse the relation between appraisal and other components of emotion from a full componential and interactional perspective. Past analyses have largely focused on features of subjective feeling as response variables. Less attention has been devoted to features of motivation, physiology, or expression. In addition, interaction hypotheses have been proposed by appraisal theories but have not been consistently studied empirically. Implementing these two assumptions should generate an empirical model that more closely matches contemporary appraisal theories.

As shown in the preceding review of the literature (Sections 1.1 and 1.2), there is currently no extant data set that is comprehensive enough to allow the type of analysis we are suggesting. In consequence, we propose to use information about emotion mechanisms contained in the semantic meaning of major emotion words in natural languages. Theories on word meaning like those found in cognitive linguistics (see Croft & Cruse, 2004, for a review) and in constructionist psychological models (e.g., Barsalou, 2008; Barrett, Wilson-Mendenhall, & Barsalou, 2015) claim that a word's meaning encompasses much of our world knowledge about the kind of thing designated by the word. For example, the meaning of "fear" would not only be "an emotion caused by the perception of threat", but would also include features pertaining to accumulated knowledge of the emotion's typical elicitation, physiological changes, expressions, and associated action tendencies. One of the potentially underlying mechanisms has been called the "lexical sedimentation hypothesis" in the classic lexical approach to personality (developed by John, Angleitner, and Ostendorf, 1988, and applied to emotion words by Fontaine et al., 2013, pp. 7-8). The central statement of this hypothesis that "Those individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely is it to become expressed as a single word." (John et al., 1988, p. 174) can be rephrased for our purposes as "Those emotion mechanisms and their outcomes that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more frequent and important a particular outcome and its underlying mechanism, the more likely is it to become expressed as a single word or category". Scherer (2005) has argued that the meaning structure of these emotion words would best be studied by domain-specific "semantic profiles" which represent the typical configuration of the major emotion component processes (see Section 1.1 above) for each word/category. We suggest that using empirically obtained semantic profiles for different emotion categories/words allows us to examine the "frozen" effects of appraisal interactions on different emotion components.

To accomplish this aim, we re-analysed the semantic profiles obtained in the large cross-cultural dataset collected by Fontaine and colleagues (2013), known as the "GRID" dataset.[3] In this study, respondents in 34 samples across 27 countries were asked to rate 24 common emotion words on their semantic (i.e., culturally shared) meaning in terms of the five components of emotion. Using a combination of dimension reduction techniques and interactional modelling, we addressed three research questions in our analysis:

1. What is the nature of the association between appraisal predictors and response features of motivation, physiology, and expression? Specifically, do we find evidence of an interactional effects structure for the appraisal predictors, as predicted by appraisal theories? No study thus far has directly investigated this issue for these three emotion components.
2. How well can we predict features of motivation, physiology, and expression from appraisal criteria? What is the overall percentage of variance that we can explain in each of these three components, and where can we find substantial gains when interactions (if any) are accounted for?
3. Can we differentiate qualitative emotion categories (e.g., anger, fear, joy) using the four-component model derived to address Questions 1 and 2? To address this question, simulated output from the obtained model was mapped onto the GRID emotion words.

Both our methodology and prior hypotheses were based on the earlier study by Meuleman and Scherer (2013). With respect to methodology, we used machine learning to estimate appropriate models for our data (multivariate adaptive regression splines, followed by nearest neighbour matching) and considered both interactions between appraisal predictors (Meuleman & Scherer, 2013) and curvilinear effects of the kind investigated by Tong and colleagues (2009). With respect to hypotheses, the earlier study pointed to a hierarchical interaction structure with a strong main effect of GOAL COMPATIBILITY appraisal and modifying effects of AGENCY and NORM COMPATIBILITY

[3] The word "GRID" refers to the semantic grid approach of the original study (see Scherer, 2005). Although it is not a proper acronym, we have chosen to adopt the same naming convention as the original authors.

appraisals on emotional responding. We therefore expected to find a similar structure in the GRID data, this time also for features of motivation, physiology, and expression as response variables. It should be noted that Meuleman and Scherer (2013) collected ratings on actual experienced emotion episodes (via self-report), whereas the GRID data represent culturally shared semantic knowledge (or "sedimented world knowledge) of emotions, as assessed by empirical surveys in 34 studies in 27 countries covering 23 languages. We therefore expect the GRID data to contain more prototypical information than the individual episodes sampled in the Meuleman & Scherer (2013) data.

2 Method

2.1 Data

Data were obtained from the GRID cross-cultural questionnaire on the meaning of emotion terms (Fontaine et al., 2013). In this study, 4948 respondents from 34 language samples in 27 countries rated the typicality of 144 items for 24 common emotion words. The 144 rated items reflected variables belonging to the five components of emotion. For example, when FEAR was the target word and the rated item was the appraisal of SUDDENNESS, the following question was asked:

> *If a speaker of your native language as spoken in your country or region uses the word FEAR to describe an emotional experience, how likely is it that the person experienced an event... that occurred suddenly?*

All items were rated in this manner on a 9-point scale ranging from "1 = extremely unlikely", over "5 = neither likely nor unlikely", to "9 = extremely likely". For our analysis, we retained 105 of the total 144 items that were pertinent to the current study.[4] This included 26 appraisal items, 40 motivation items, 18 physiology items, and 21 expression items. All items related to the feeling

[4] The 39 omitted items represented 5 appraisal items about supernatural agency and core-relational themes, 5 expression items that overlapped conceptually with some of the motivation items, all 20 feeling items, 2 items about duration and intensity, 4 items about emotion regulation, and 3 items about miscellaneous cultural aspects of emotion.

component were dropped, as we wished to focus explicitly on features of motivation, physiology, and expression as the response variables in our analysis. The ratings for the retained items were first standardized within subjects to remove response bias, and then averaged by sample by emotion word (34 × 24), yielding 816 cases. For more information on the distribution of these data by component and across cultures, as well as the psychometric qualities of the GRID instrument, we refer to the original study by Fontaine et al. (2013).

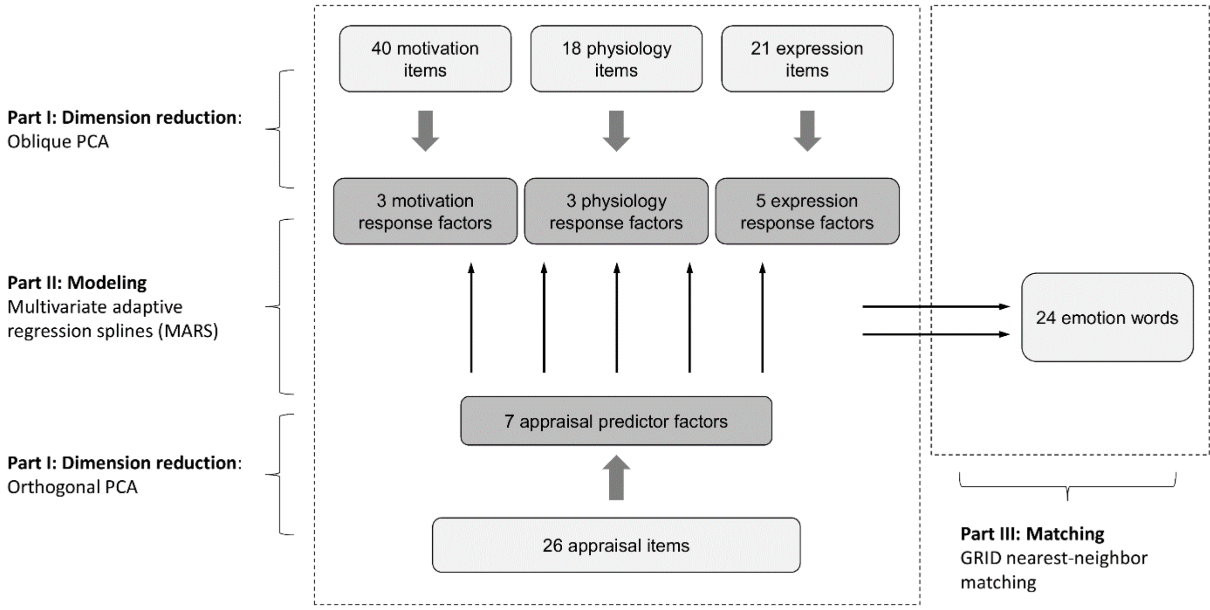

Figure 2. Modelling strategy for the GRID data. In the first part, both appraisal items and items for motivation, physiology, and expression are factorized with PCA. In the second part, the motivation, physiology, and expression factors are modelled simultaneously using multivariate adaptive regression splines (MARS). In the third part, the output of the MARS model is matched to the 24 emotion words using nearest-neighbour matching.

2.2 Data Analysis

The data analysis proceeded in three parts, (i) dimension reduction, (ii) modelling, and (iii) matching (see Figure 2).

Dimension reduction. We used PCA to reduce each set of GRID items into a smaller number of factors for each emotion component.[5] We closely followed the procedure of Fontaine and colleagues (2013), reducing the 26 appraisal items to 7 appraisal factors, the 40 motivation items to 3 motivation factors, the 18 physiology items to 3 physiology factors, and the 21 expression items to 5 expression factors (Figure 2). The only thing we changed in the present study was the rotation procedure of the PCA. For the appraisal predictor factors, we retained the orthogonal varimax rotation of the original analysis. This allowed us to identify the unique contribution of each factor during the modelling stage, and to stabilize our modelling algorithm, which is known to be sensitive to correlations among predictor variables (see next section). For the motivation, physiology, and expression response factors, however, we used an oblique promax rotation (with default rotation power, $m = 4$). The calculation of correlations among response variables rather than predictor variables is desirable from a multivariate modelling perspective. This is because when the correlation among the response variables is higher, it makes more sense for them to share the same predictor set. Using a promax rotation instead of a varimax rotation had a negligible impact on the interpretation of the motivation and expression response factors, but did affect the interpretation of the physiology response factors. Unlike Fontaine et al. (2013), who obtained a DISTRESS SYMPTOMS, AROUSAL, and TEMPERATURE factor, we obtained a RELAXATION VS TENSION, AROUSAL, and TEMPERATURE factor (see results section for interpretation and supplementary material, Table S1, for the full loading matrix). Note, finally, that all factor scores were computed using Thurstone's method.

Modelling. We simultaneously modelled the 11 response factors as a function of the 7 appraisal predictor factors, using *multivariate adaptive regression splines* (MARS; Friedman, 1991). This part of the analysis addressed the first and second of our three research questions. The MARS model is a variant of the standard linear regression model. As in ordinary multiple regression, we model one or more response variables as a function of one or more predictor variables. The MARS model increases the flexibility of the standard linear model in two important ways: first, by

[5] In this paper, we use the term factor for principal components to avoid confusion with components of emotion

constructing a piecewise linear function rather than a single straight line/plane; second, by automatically selecting relevant interactions between predictor variables.

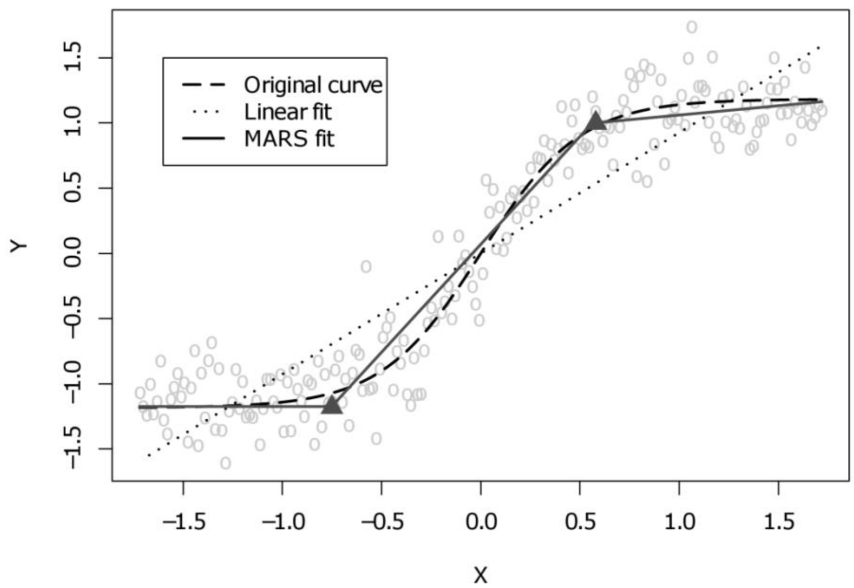

Figure 3. Comparison of a linear fit versus a MARS fit for a nonlinear curve. The MARS fit achieves more flexibility by constructing a piecewise linear function. The black triangles depict the hinges of each spline.

The first feature is illustrated in Figure 3. A standard linear model (dotted line) and the MARS model (solid line) attempt to approximate some curvilinear data (dashed line). The MARS model achieves a better approximation of this curve than the standard linear model by constructing a piecewise linear slope. The pieces of this function are known as splines and are joined by hinges (i.e., knot points). MARS will automatically select the optimal number of hinges, their location, and the number of splines. It is more “sparse” than standard regression in the sense that MARS only models those areas of the data where it can find significant associations, and sets the response to a constant elsewhere. The major advantage of the spline approach is that we achieve a flexibility similar to non-parametric curves (e.g., LOWESS smoothers) while retaining easily interpretable linear slopes.

MARS selects the optimal hinge points by a stepwise selection that minimizes the generalized cross-validation error of the model (*GCVE*). The *GCVE* criterion is a so-called penalized measure of fit, like *AIC* or *BIC*, and hence robust to overfitting. The selection of main effects and interaction effects occurs likewise. MARS only adds an effect to the model if it lowers the *GCVE* criterion.[6] This procedure automatically selects the maximal degree of interaction (e.g., 2nd order, 3rd order). When more than one response variable is modelled simultaneously (i.e., a multivariate regression), MARS only adds hinges and effects if they reduce GCVE *across all response* variables. The final result of the selection process is a model with piecewise linear terms and interactions that can be interpreted in the same fashion as an ordinary linear model. It is important to note, however, that the MARS procedure completely avoids classical hypothesis testing. The final model does not return *t*-tests or *p*-values, which would be redundant for two reasons. First, in our data specifically, all relations would have turned out to be significant due to excess statistical power ($N = 816$). Second, the selection of an effect by MARS is indicative of predictive relevance in itself. If a predictor has no meaningful association with the set of response variables, MARS does not select this variable for inclusion in the final model. The reduction in *GCVE* is a measure of predictive strength, and not sensitive to overfitting, excess power, or issues of multiple testing, making this criterion superior to classical significance testing.

A potential drawback of the MARS model is that it is unstable in the presence of correlated predictor variables (Milborrow, 2011a). This was an important reason for us to favor orthogonal appraisal factors—as computed by our PCA procedure—rather than using the original set of 26 appraisal items or allowing oblique rotation among appraisal factors. In addition, it allowed us to quantify variable importance free of confounding.

To evaluate the MARS procedure for our data, we quantified the relative contribution of splines and interactions to the model fit by systematically removing these options from MARS and evaluating the reduction in cross-validated R^2, which is a measure of explained variance based on a 20-fold cross-validation of the data. This measure also served as our primary index of predictive strength for the final model. To aid interpretation, we also computed estimates of relative variable importance and surface plots to visualize interactions.

---

[6] Additionally, adding interactions is subject to the constraint that at least one involved variable has been retained as a main effect.

Simulation and matching. In the final part of the analysis, simulated output from the estimated MARS model was matched to the emotion words of the GRID data. This part addressed the third research question. We first generated model-based predictions for interesting combinations of appraisal values, and then matched the combined predictor and fitted values to the nearest similar case in the GRID data, using a nearest-neighbour algorithm (Euclidean distance). This procedure allowed us to connect values of appraisal, motivation, physiology, and expression features simultaneously to emotion words in the manner prescribed by the componential perspective on emotion (Figure 1). Consistent with a previous study (Scherer & Meuleman, 2012), a nearest-neighbour approach was chosen because of its relatively non-parametric nature. The matching analysis allowed us to generate maps of emotion labels corresponding to particular combinations of values from the four-emotion-component MARS model. We analyzed these maps primarily in a qualitative rather than a quantitative manner because this analysis depended on novel simulations, and these cannot be evaluated against a corresponding "ground truth" emotion category. However, we did assess how the certainty of these predictions shifted across emotion categories and across the appraisal criteria we manipulated for the simulation.

2.3 Software

All analyses were carried out using the R statistical software, version 3.0.2. Principal component analysis was conducted with the "psych" package, version 1.3.10 (Revelle, 2013). The analysis of multivariate adaptive regression splines was conducted with the "earth" package, version 3.2.6 (Milborrow, 2013b). The random effects ANOVA was conducted with the "lme4" package, version 1.0.5 (Bates et al., 2013) and the "lmerTest" package, version 2.0.0 (Kuznetsova, Brockhoff & Christensen, 2013). Nearest-neighbour matching was conducted with the "kknn" package, version 1.2.5 (Schliep & Hechenbichler, 2014), and "ipred" package, version 0.9.2 (Peters & Hothorn, 2013), for cross-validation and prediction, respectively. Three-dimensional surface plotting and correlation plotting were conducted using the "plotmo" package, version 1.3.2 (Milborrow, 2012b) and the "corrplot" package, version 0.73 (Wei, 2013), respectively.

3 Results

3.1 Dimension Reduction

The names of the 18 factors that were extracted from the original GRID items and their correlations are presented in Figure 4 (see Fontaine et al, 2013, for a full description of the dimension reduction procedure). Most of the factors are straightforward (e.g., jaw drop). We briefly zoom in on those factors that require some additional clarification. The appraisal factors GOAL COMPATIBILITY, AGENCY, and NORM COMPATIBILITY are bipolar factors, with values ranging from goal incompatible to goal compatible, self-caused to other-caused, and norm incompatible to norm compatible, respectively. The motivational factors are all bipolar and reflect tendencies to act in a certain way. The APPETITIVE VS DEFENSIVE factor corresponds to the tendencies to approach/continue (positive values) versus retreat/stop (negative values). The INTERVENTION VS DISENGAGEMENT factor corresponds to tendencies to undertake any action or movement (positive values) versus to remain still/freeze (negative values). The ATTACK VS SUBMIT factor corresponds to tendencies to dominate (positive values) versus *to be* dominated (negative values). The AROUSAL factor refers to general symptoms of physiological arousal (e.g., increased heartbeat, blushing, shivers), TEMPERATURE refers to the felt body temperature (positive values indicating higher, negative values indicating lower temperature), and RELAXATION VS TENSION refers to the relaxing (positive values) versus tensioning (negative values) of the muscles. The FROWN VS SMILE factor is a bipolar factor reflecting frowning versus (positive values) smiling (negative values) expression, and VOCAL ENERGY refers to volume and speed characteristics of the voice.

Figure 4 depicts the bivariate Pearson correlation between each pair of factors. In this matrix, correlations with values smaller than 0.30 were omitted from plotting, following Cohen’s (1988) rule of thumb for small effect sizes. The correlation matrix in Figure 4 reveals that factors were relatively uncorrelated *within* emotion components. For the appraisal component, this was imposed by the varimax rotation. For the other emotion components, where oblique rotation methods were applied, correlations were still relatively small. More substantial correlations were found *between* emotion components, pointing to linear coherence in emotional responding across components. Interestingly, the majority of the appraisal predictor factors did not correlate substantially with most of the response factors.

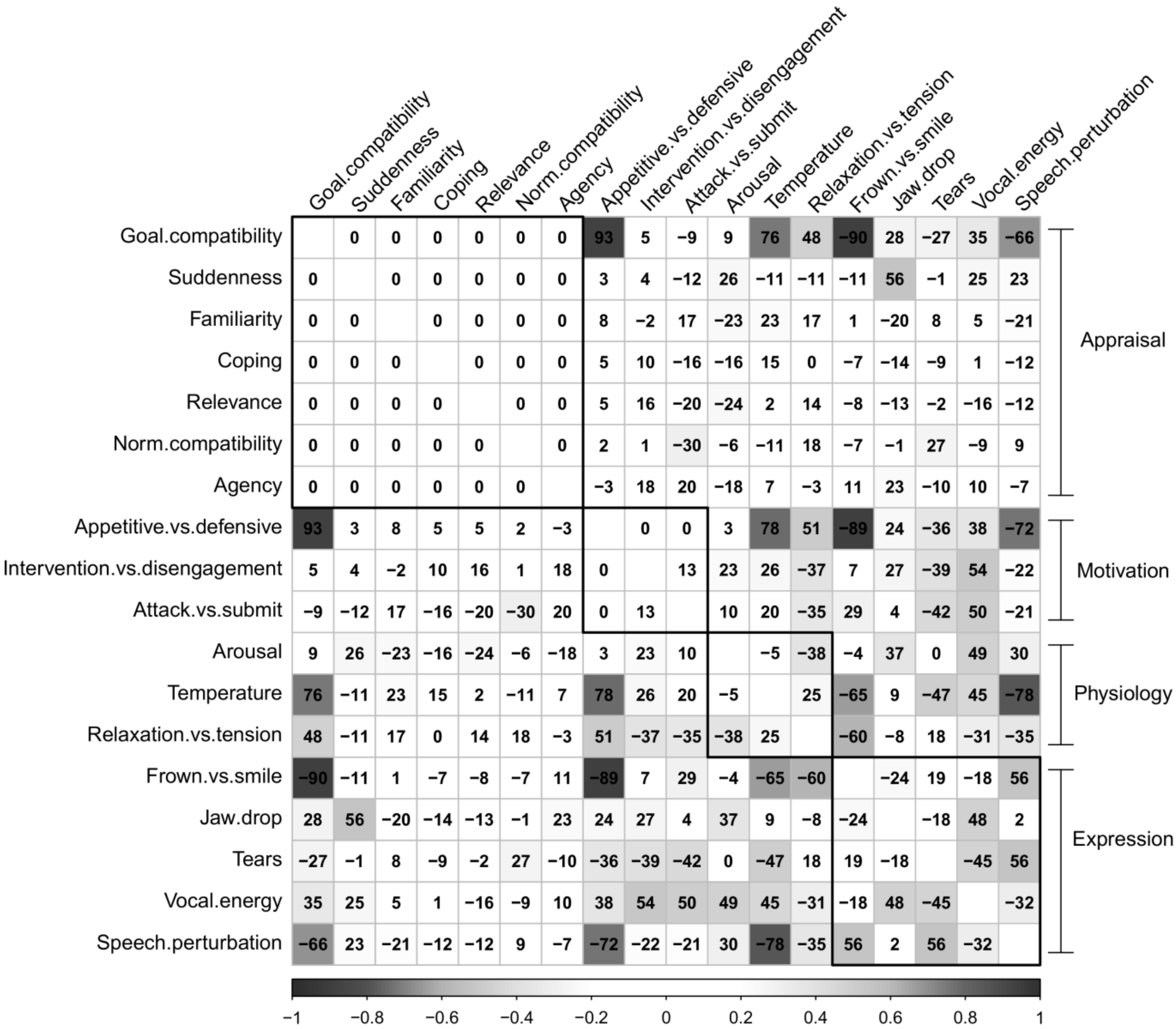

Figure 4. Factors extracted from the GRID data and their correlations multiplied by 100. Diagonal elements have been removed for visual clarity. For bipolar factors, the label of the positive end of the scale is named first, e.g., attack (positive) versus submit (negative).

3.2 Modelling

The MARS algorithm was applied using the 7 appraisal factors as predictor variables and the 11 motivation, physiology, and expression factors as simultaneous response variables. MARS selected a 3rd-degree interactional structure as the final model, including spline effects. Table 1 gives an overview of the selected appraisal factors and the degree of interaction at which they appear. All appraisal factors appeared in the final model, either as a main effect or as part of interactions. In other words, no appraisal factor was found to be redundant for predicting the 11 component response factors. Among these, the GOAL COMPATIBILITY factor was the single most important predictor of component responses, appearing at all three levels of interaction (main, 2nd, 3rd) and achieving the largest reduction in *GCVE* by MARS (Table 1, relative importance column). The appraisal factors of FAMILIARITY and COPING uniquely showed main effects on emotional responding, and did not interact with other appraisal factors. The factors of AGENCY and NORM COMPATIBILITY were found to be uniquely involved in interactions (2nd and 3rd degree), indicating that these were the most important *modifiers* in the model. Already, these results paint a different picture than the linear bivariate correlations in Figure 4. Finally, Table 1 lists the number of hinges that were used for constructing the splines per appraisal predictor factor. Only GOAL COMPATIBILITY required more than one hinge for changing the direction of the slope.

Table 1. Final selected MARS model

	Degree of interaction				
Appraisal factor	Main	2nd	3rd	Rel. importance	Hinges
GOAL COMPATIBILITY	Yes	Yes	Yes	100.0	4
SUDDENNESS	Yes	Yes		64.5	1
FAMILIARITY	Yes			35.0	1
COPING	Yes			30.7	1
RELEVANCE	Yes	Yes		39.1	1
NORM COMPATIBILITY			Yes	56.0	1
AGENCY		Yes	Yes	56.0	1

Degree of interaction, relative importance, and number of spline hinges for appraisal predictors in the final selected MARS model. Variable relevance is calculated relative to the most important variable, which is set to 100 by default.

To what extent did the use of splines and interactions contribute to the overall model fit? We manipulated the presence or absence of these features and inspected the change in cross-validated R^2 across 20 cross-validation folds in the data and for each response factor, yielding 880 R^2_{cv} values.

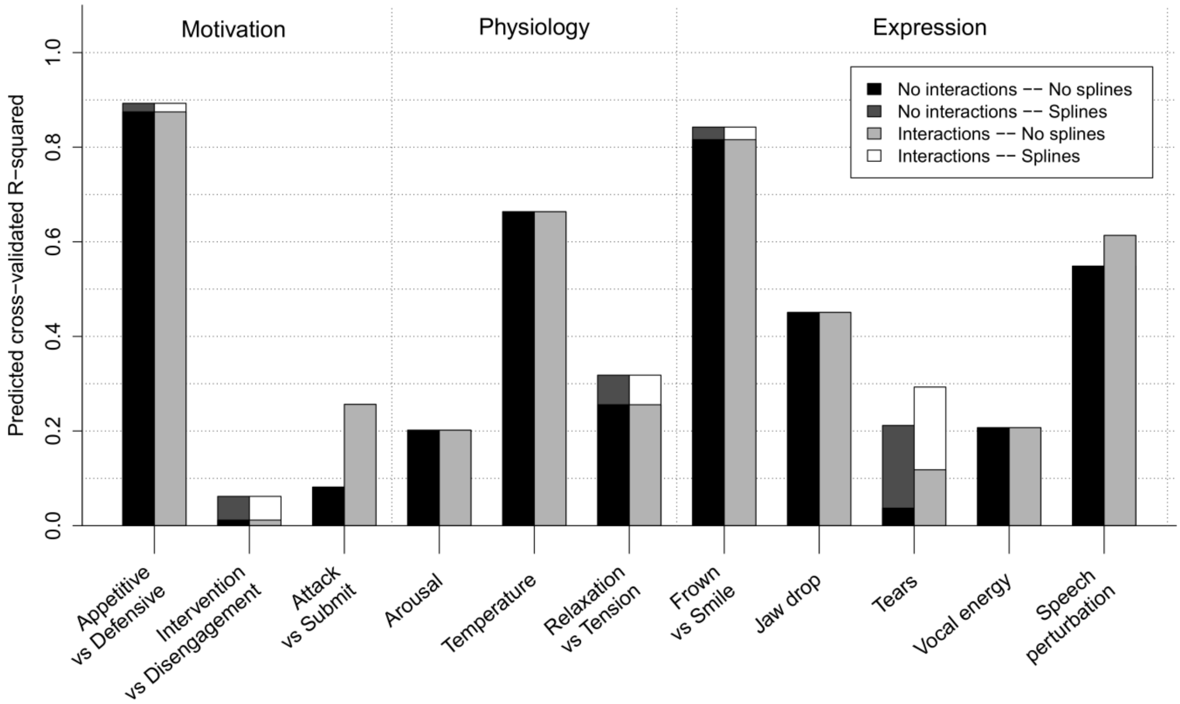

Figure 5. Estimated changes in cross-validated R^2 as a function of adding interactions and/or splines to the MARS model. Adjacent bars of different height indicate an interaction effect. Stacked bars indicate a spline effect (e.g., AROUSAL has neither effect, whereas TEARS has both effects).

These values were then submitted to a 2 × 2 × 11 random effects ANOVA of SPLINES (with or without) by INTERACTIONS (with or without) by RESPONSE (one of 11 response factors). The cross-validation fold was treated as a random intercept. Denominator degrees of freedom for all *F*-tests were adjusted using Satterthwaite's approximation for random effects models (Fitzmaurice, Laird & Ware, 2004). No evidence was found for a three-way interaction between RESPONSE, SPLINES, and INTERACTIONS, $F(10,816.3) = 0.257$, $p = 0.990$, nor for a two-way interaction between SPLINES and INTERACTIONS, $F(1,826.28) = 0.327$, $p = 0.567$. We did find evidence of two-way interactions between RESPONSE and SPLINES, $F(10,827.28) = 2.829$, $p = 0.002$, and between RESPONSE and INTERACTIONS, $F(10,827.28) = 3.543$, $p < 0.001$. In other words, adding splines or interactions improved R^2_{cv} differently between factors. To investigate these interaction effects, we refitted the random effects model for each response factor separately, using only additive effects for SPLINES and INTERACTIONS.[7] Results of this analysis are reported in Figure 5.

From Figure 5, we can draw several conclusions. First, the motivation, physiology, and expression components were, on average, predicted equally well by the appraisal predictor factors in the MARS model, with mean R^2_{cv} equal to 42.0%, 41.6%, and 48.7%, respectively. Second, sizeable differences in explained variance were found between individual response factors, with R^2_{cv} reaching as much as 89.0% for the APPETITIVE VS DEFENSIVE motivation factor, and as little as 7.3% for the INTERVENTION VS DISENGAGEMENT factor. Six out of eleven response factors reached an R^2_{cv} larger than 30%, indicating that overall predictive power was satisfactory. Third, evidence for nonlinear effects was substantial, supporting theoretical claims for interactions and curvilinear patterns. Seven out of eleven response factors showed a significant improvement when adding either splines or interactions to the MARS model. The remaining four factors, AROUSAL, TEMPERATURE, JAW DROP, and VOCAL ENERGY, showed no improvement in R^2_{cv} from adding such features (equal bars without stacked area in Figure 5), indicating that these factors could be modelled equally well with a standard linear appraisal model. Overall, nonlinearity was the least substantial for the physiology component. Fourth, some response factors improved on R^2_{cv} when adding splines but not when adding interactions, namely APPETITIVE VS DEFENSIVE, INTERVENTION VS DISENGAGEMENT, RELAXATION VS TENSION, and FROWN VS SMILE. Not surprisingly, these factors are all bipolar, indicating that trends differed between the negative and positive ends of these scales. Finally, some factors benefitted from interactions but not splines, namely ATTACK VS SUBMIT and SPEECH PERTURBATION, indicating that these factors of emotional responding crucially depended on modifications of AGENCY and NORM COMPATIBILITY appraisal (see Table 1 for the general interaction structure). Finally, one response factor, TEARS, showed both substantial spline *and* interaction effects, improving R^2_{cv} by over 20%.

[7] We used an additive model because the overall ANOVA showed no SPLINES by INTERACTIONS effect.

The full coefficient table of the MARS model would be too large to report and too complicated to interpret (see supplementary material, Table S2). Instead, we can visualize the spline and interaction effects for selected combinations of appraisal predictor factors and response factors, using line and surface plots. For response factors that did not benefit from interaction effects (Figure 5) it suffices to visualize one appraisal predictor at a time. Figure 6 shows three examples: The line plots visualize the predicted rate of change for three different response factors as a function of three different appraisal predictors when the non-visualized appraisal predictors are set to their mean value of 0. These plots also visualize the hinges that were selected for constructing the splines.

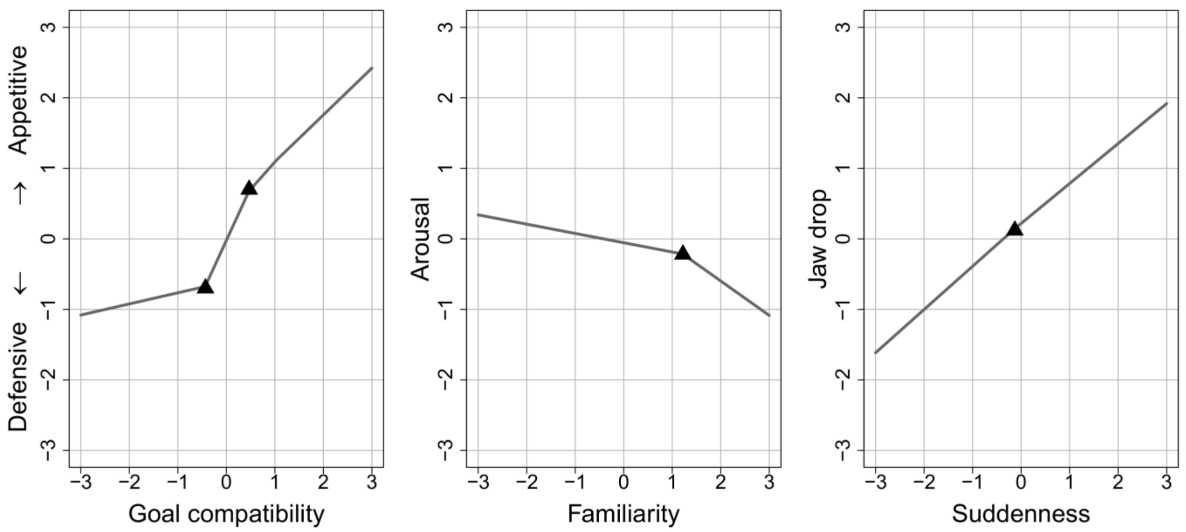

Figure 2. Effects of three appraisal predictor factors on three response factors, when all other appraisal predictor factors are fixed to their mean value of 0. All factors on a standardized factor scale with mean 0 and standard deviation 1.

For the relation between GOAL COMPATIBILITY and APPETITIVE VS DEFENSIVE tendency (Figure 6, left panel), we see that the rate of change differs for both ends of the factor scale. Overall, we see that increasing appraisal of incompatibility increases defensive action tendencies, whereas increasing appraisal of compatibility increases appetitive action tendencies. For the latter, the rate of change is much steeper than for the former. For the relation between FAMILIARITY and AROUSAL (Figure 6, middle panel), increasing familiarity simply leads to decreasing physiological arousal. This decrease occurs even faster for strong appraisal of familiarity. For the relation between SUDDENNESS and JAW DROP (Figure 6, right panel), we see that the more sudden an event is appraised to be, the more jaw drop expression is reported. For this relation, the spline effect hardly matters, as the rate of change remains more or less identical before and after the hinge.

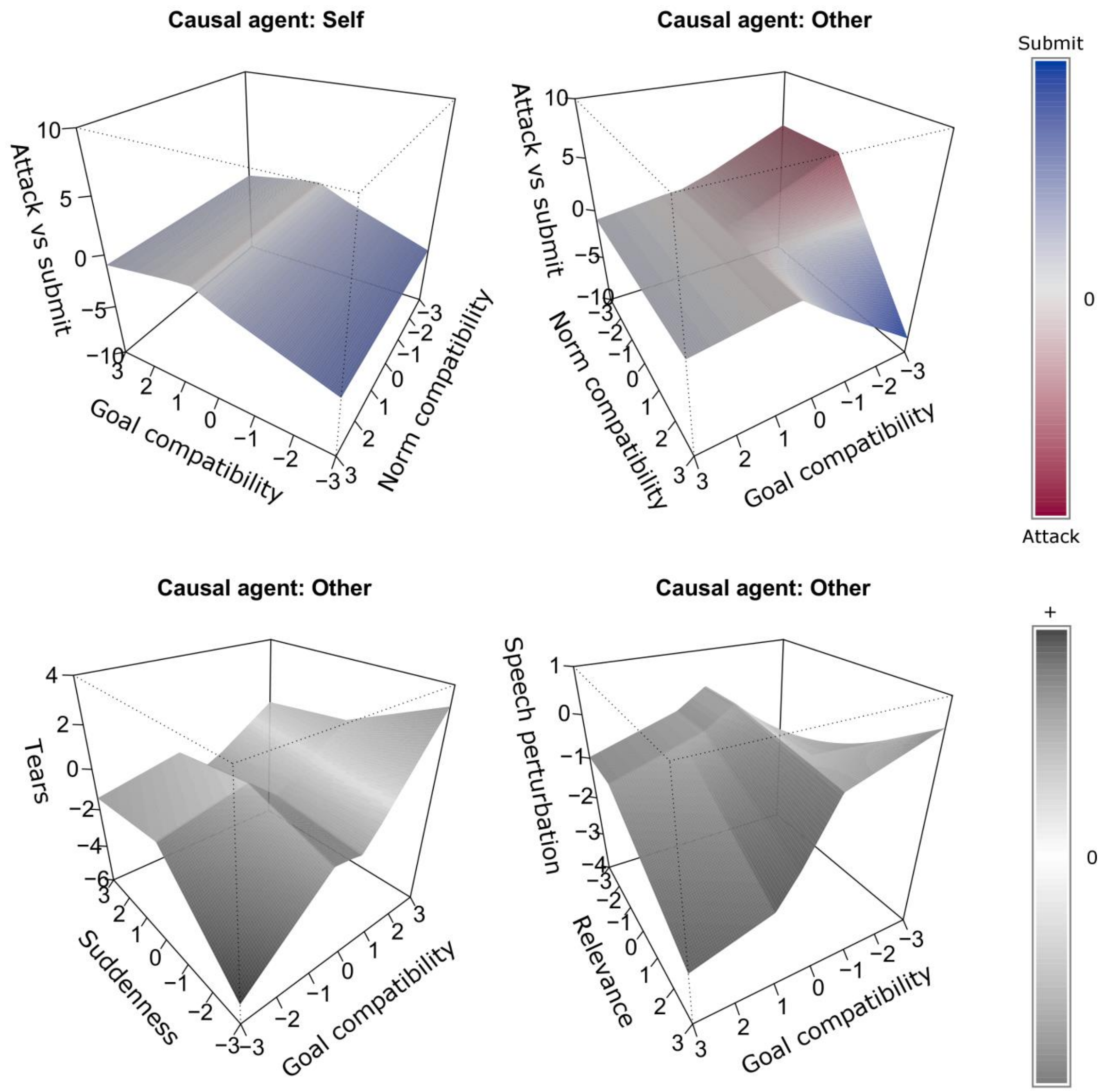

Figure 7. Three-way interaction between goal compatibility, norm compatibility, and agency appraisal. For this plot, appraisal of agency has been dichotomized into other agency (-3) versus self-agency (+3). Surfaces are plotted for the approach–reject and assert–submit response factors. All factors on a standardized factor scale with mean 0 and standard deviation 1.

For response factors that benefited from interaction effects (Figure 5), we need to visualize several appraisal predictor factors simultaneously. This produces three-dimensional surface plots of the kind depicted in Figure 7. The top two panels of this figure visualize the three-way interaction of GOAL COMPATIBILITY × NORM COMPATIBILITY × AGENCY on the ATTACK VS SUBMIT tendency. Shades of blue indicate a stronger submit tendency, whereas shades of red indicate a stronger attack tendency. When the person appraises herself as the cause of an event (top left panel), no attack tendency is reported for any combination of appraised goal and norm compatibility, only a submit tendency. In addition, the degree of submit tendency does not depend on appraisal of norm compatibility, but simply on appraisal of goal compatibility: More incompatible events lead to a stronger submit tendency. When someone else is appraised as the cause of an event (top right panel), the response surface changes drastically. This time, an attack tendency is predicted for combinations of high goal incompatibility and high norm incompatibility. For events appraised as goal incompatible yet norm compatible, the model once again predicts a submit tendency.

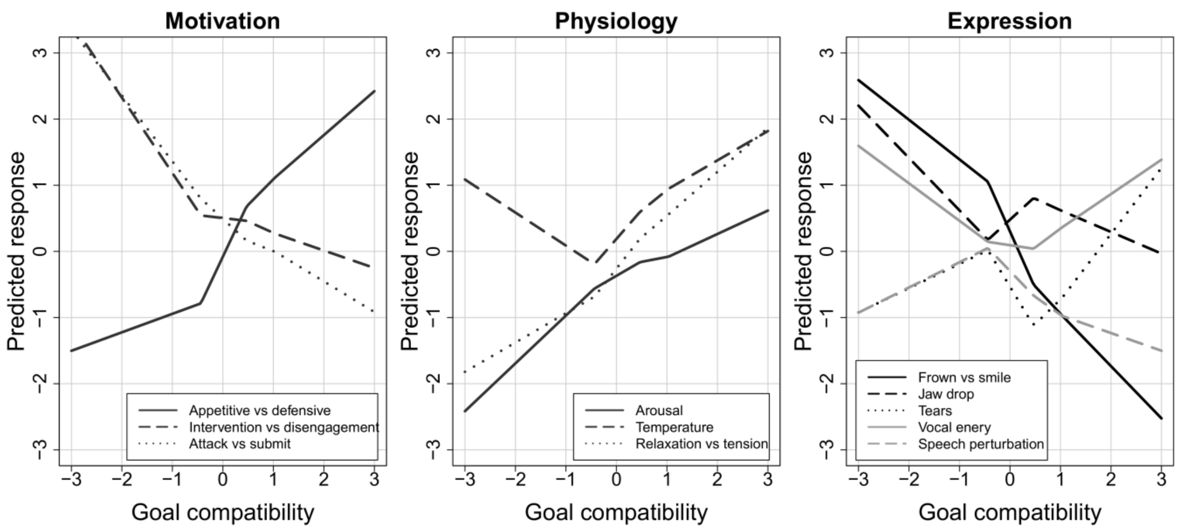

Figure 8. Multicomponential response pattern as a function of goal compatibility appraisal. Note that the appraisal of agency was fixed to other-agency (+3), and all other appraisal factors were fixed to their mean value of 0. All factors on a standardized factor scale with mean 0 and standard deviation 1.

The bottom two panels of Figure 7 visualize the GOAL COMPATIBILITY × SUDDENNESS and GOAL COMPATIBILITY × RELEVANCE interactions for the response factors TEARS and

SPEECH PERTURBATION, respectively. These effects were conditioned on another person being appraised as the cause of an event. Tears become less likely for decreasing appraisal of suddenness but this trend reverses when an event is appraised as goal compatible. In fact, the highest tears response is predicted for appraisals of low suddenness and high goal compatibility. A similar reversal can be observed in the response surface for speech perturbation. Speech perturbation increases with increasing goal incompatibility, except when an event is appraised as being totally irrelevant.

Finally, we can also visualize the complete multicomponential response to the appraisal of goal compatibility, which gives a better idea of the full emotional state that is being predicted. This visualization is depicted in Figure 8. In this plot, goal compatibility was allowed to vary, appraisal of agency was set to other-agency, and all other appraisal predictors were set to their mean value of 0. This plot shows how increasing the appraisal of goal incompatibility produces a multicomponential response that is indicative of anger, including increasing intervention and attack motivation, increasing body temperature, increasing frowning, and increasing vocal energy.

3.3 Simulation and Matching

So far, we analysed emotions in terms of appraisal, motivation, physiology, and expression. In the final part of the analysis, we wanted to reintroduce the emotion words that were part of the GRID database and connect these to the estimated MARS model, using a nearest-neighbours algorithm.

First we simulated new fitted values of the MARS model based on inputted combinations of GOAL COMPATIBILITY, NORM COMPATIBILITY, and AGENCY values. Although other combinations of appraisal values could have be chosen, we decided to focus on the three appraisal factors involved in the three-way interaction of the MARS model. The former two appraisal factors were allowed to vary continuously from -3 to 3, in increments of 0.05, while the latter appraisal factor was fixed to either -3 (self-agency) or +3 (other-agency). All other appraisal factors were fixed to their average value of 0. These values were manipulated orthogonally to produce a total of 29,282 appraisal combinations, which were then inputted into the MARS model to simulate new fitted values in the response components. For each such simulation, this produced a combined pattern of appraisal input and fitted response values. An example pattern is depicted in Figure 9 (top panel). Then, each of these combined patterns was matched to the nearest similar cases in the observed GRID data by its Euclidean distance (see Figure 9, top panel, for a comparison of the simulated pattern with its nearest neighbor).[8] The majority emotion word among these nearest cases in the observed GRID data was then chosen as a "predicted label" for the simulated pattern. This resulted in the map of matched labels that is depicted in Figure 9 (bottom panels).

Figure 9 shows the matched emotion labels form a structured geography. Appraisal of GOAL COMPATIBILITY linearly separates positive from negative emotions. That is, events appraised as goal compatible generate emotional responding that is most proximal to labels such as pride, pleasure, contentment, happiness, and joy. Events appraised as goal incompatible, on the other hand, generate emotional responding that is most proximal to labels such as guilt, anger, disgust, jealousy, hate, disappointment, and sadness. Figure 9 also highlights the effect of agency appraisal, which produces a dramatically different emotion landscape depending on whether the person takes herself (left panel) or another person (right panel) as the cause of an event. This picture is made more complex in the presence of norm violations. For example, for self-agency, an appraisal of goal incompatibility generates emotional responding that is close to guilt, irrespective of an appraisal of norm violation. By contrast, an appraisal of goal compatibility generates emotional responding that is close to pride *only* when the event is also appraised as norm compatible. When an event is appraised as norm incompatible, pleasure becomes more likely.

[8] Leave-one-out cross-validation showed that a parameter of k=13 neighbours achieved the lowest misclassification error (26.5%) for the factorized GRID data.

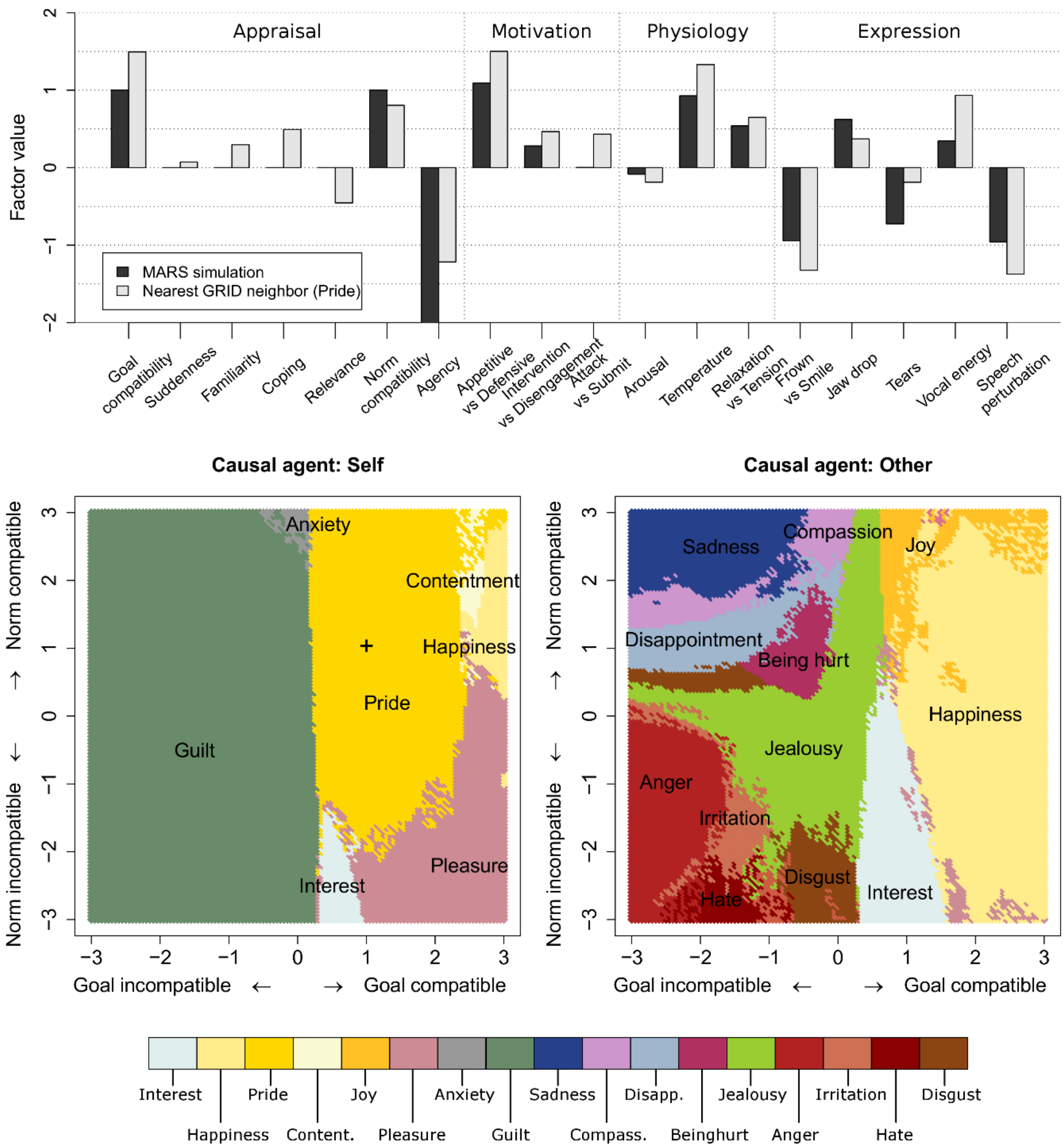

Figure 9. Simulation and matching analysis of the GRID data. Top panel: Simulated pattern of data based on appraisal input to the fitted MARS model (dark bars), juxtaposed with the nearest neighbor of the observed GRID data (light bars). Bottom panels: Map of simulated data patterns matched to the emotion label of the nearest observed GRID cases, indexed by varying appraisals of goal compatibility, norm compatibility, and agency. The remaining appraisal factors were fixed to their average value of 0. All factors on a standardized factor scale with mean 0 and standard deviation 1. The pattern marked with a black cross corresponds to the simulation depicted in the top panel of this figure.

The landscapes in Figure 9 conform largely to intuitions about the link between the three manipulated appraisal criteria, the associated component response patterns, and the matched emotion label. Evaluating the goodness-of-fit of this landscape in a quantitative manner is complicated due to the fact that the simulated patterns possess no ground truth label. Novel predictions generated by the MARS model can therefore only be validated in follow-up studies. Nevertheless, we performed additional analyses of the nearest neighbor predictions by assessing how the certainty of these predictions shifted across emotion categories and across the manipulated appraisal criteria (Table S3; Figure S1). As a measure of probability, we looked at the proportion of the majority label among the 13 nearest neighbors (e.g., if 8 out 13 neighbors are "pride", the probability of "pride" equals $8/13 \approx 61.5\%$). Across all emotions, the average probability of a prediction was 41.8%, with the highest average probability for guilt and sadness predictions (68%) and the lowest for anxiety (23%; Table S3). This was much higher than chance level, at 4.2%. Next we fitted a regression model with probability of a nearest neighbor prediction as the dependent and the manipulated appraisal criteria as the independents. Results indicated that, for self-agency simulations, higher values of goal compatibility led to more uncertain nearest neighbor matching. For other-agency simulations, combinations of low goal compatibility and low norm compatibility led to more uncertain nearest neighbor matching (Figure S1). For no simulated scenario did the prediction probability drop below chance level, however.[9] These results suggest that the structure of emotion labels observed in Figure 9 was not random.

4 Discussion

Appraisal theories assume that emotions are multicomponential phenomena, comprising changes in appraisal, motivation, physiology, expression, and feeling. In addition, many appraisal theories assume that appraisal criteria interact to differentiate emotions. This study operationalized these two assumptions in a single model—multivariate adaptive regression splines—and tested them in a large dataset on culturally shared semantic profiles of emotion words (GRID data; Fontaine et al., 2013). We addressed three research questions, which we now discuss in light of the present results.

[9] Note that, due to the large amount of data in this model, inferential tests were overpowered and therefore not meaningful to interpret (all significant).

Interactions and nonlinearity. Our first research question concerned the association between appraisal criteria and features of motivation, physiology, and expression. While many appraisal theories have put forward interactional and even curvilinear hypotheses, these have not been systematically investigated or they have been limited to using feelings as response variables. In our study, we found substantial evidence for interaction effects and curvilinear associations across features of motivation, physiology, and expression. The final MARS model included both splines and interactions up to the third degree. Appraisal of goal compatibility played a central role in this model, appearing as a main effect and in all higher-order interactions. The prominence of this appraisal criterion is due to its power in differentiating positive from negative emotional responses (see Figure 9). More subtle emotion differentiation tended to occur via interactions with the other appraisal criteria, such as appraisal of suddenness, relevance, agency, and norm compatibility. Interestingly, the latter two criteria appeared exclusively in interactions, without having a corresponding main effect. These results converge with the interaction structure that was obtained by Meuleman and Scherer (2013) for recalled emotion experiences. In that study, appraisal of goal compatibility also operated as a strong main effect, whereas agency and norm-related appraisal criteria appeared only in interaction terms (see also Tong et al., 2005, for a pure interaction effect of moral violation). These results validated the interaction assumptions of appraisal theories. Nevertheless, the strong hierarchical interactivity proposed by Roseman (2001) and Scherer (2001) could not be substantiated. No evidence was found for a single main effect characterized by subsequent pure interactions. Instead, models were found to be partly additive and partly hierarchical. Interaction effects beyond the third degree were not selected by the final MARS model.

Our application of splines extended the work of Tong and colleagues (2009) on curvilinear associations between appraisal criteria and features of feelings, and allowed our model to adapt to local patterns in the data. In some cases the spline effects suggested that an appraisal criterion must reach a certain threshold value before it can affect emotional responding (see Figure 6, middle panel). In other cases the spline effects allowed a slope to completely reverse its direction before and after the hinge point (see Figure 8, middle panel).

Although the final MARS model included both splines and interactions, a follow-up analysis showed that not all response factors benefitted equally from these features. The factors AROUSAL, TEMPERATURE, JAW DROP, and VOCAL ENERGY were predicted equally well by a linear additive model. This may be due to the fact that variance in each of these response factors can be largely accounted for by a single appraisal predictor. For example, appraisal of suddenness predicted the degree of jaw drop almost completely on its own (Figure 6, right panel). Splines were mainly beneficial for modelling bipolar response factors (e.g., FROWN VS SMILE). This suggests that different poles of such factors require different modelling. Finally, interactions were the most beneficial for modelling ATTACK VS SUBMIT, TEARS, and SPEECH PERTURBATION, suggesting that a correct response differentiation in these factors depends strongly on appraisals of agency and norm compatibility (see Figure 7, top panels).

Predicting motivation, physiology, and expression. Our second research question concerned the predictive strength of appraisal on features of motivation, physiology and expression. Overall, the amount of explained variance achieved by the final MARS model was substantial, with R_{cv}^2 as high as 89.0% for the APPETITIVE VS DEFENSIVE tendency. Although individual response features differed with respect to the amount of variance explained by appraisal criteria (Figure 5), the three emotion components were predicted equally well on average (between 40% and 50% of the variance). These results pointed to the relatively high interdependence among emotion components in the GRID data, and suggests that each of these components plays a significant role in determining the final content of the emotion.

Differentiating emotion words. Our third research question concerned whether we could cluster and label emotions (e.g., joy, anger, fear) that were simulated from the four-component model that was derived during modelling. To do so, we manipulated values of GOAL COMPATIBILITY, NORM COMPATIBILITY, and AGENCY to generate new output (i.e., fitted values) in the MARS model. The new output was then labelled with a feeling term by matching it to the most similar cases from the observed GRID data (using Euclidean distance nearest neighbour matching). This resulted in a map of emotion words that varied according to the manipulated appraisal criteria (Figure 9), and which indictaed the linear separation of positively and negatively valenced emotions by appraisal of goal compatibility (see Meuleman and Scherer, 2013, for a highly similar map). Within those two valence regions, subtle emotion differentiation occurred along the norm compatibility and agency appraisals. Since our matching procedure was an unsupervised method of classification, we could not compare the predicted labels directly to ground truth labels. However, the structure of the landscape could be assessed and individual label predictions could be evaluated for their plausibility in matching the underlying pattern of values in the four emotion components. For example, in Figure 9, the simulated pattern marked with a cross on the bottom left landscape was matched to “pride”. The barplot in the top panel of this figure shows that the underlying pattern was generated from high goal compatibility appraisal, high norm compatibility appraisal, and self-agency appraisal. Prominent component responses include high appetitive tendencies, high temperature increases, smiling, and lowered speech perturbation.

This example highlights the utility of our approach. Contrary to most studies on emotion, we decided not to focus directly on the feeling component as the main phenomenon to be explained. Instead, we focused on predicting aspects of motivation, physiology, and expression. Labelling with emotion words occurred only at a late stage of the analysis and in an unsupervised manner. In doing so, we recall that labelled states (a) can be a fuzzy set of categories open to interpretation and relabeling, and (b) may communicate less about the meaning of an emotional state than commonly assumed. In the final MARS model that we derived, emotional output was generated unambiguously in terms of potential or manifested *behaviours and actions*. An emotion label would not add much information to this response pattern except as an integrated summary, much in the same way that we use these labels in day-to-day life to communicate our feelings.

As an application, consider the problem of creating an affective agent such as a video game character or a robot imbued with emotions. Now consider two candidate appraisal models for this problem, one that outputs emotion labels directly (e.g., OCC model), and our MARS model that only outputs the label optionally. In the first case, the agent would not be able to do much more than communicate the label that was generated (e.g., *“I am angry”* following the appraisal of a goal incompatible event). In the second case, the agent would display changes in motivation and bodily symptoms following appraisal (e.g., attack tendency, blushing, frowning). The anger pattern would be clear when observing the agent, even without adding the label explicitly to it.

Limitations and future work. It was highlighted in the introduction that the data, the analyses and the interpretations reported in the current article were obtained on the basis of empirical assessments of culturally shared semantic profiles of major emotion words (reflecting sedimented world knowledge) rather than on self-reported memories of actual emotion situations experienced by individuals. We believe that this should not be considered a limitation in the sense that it limits the validity of our data as we do not claim to have examined actual, day-to-day emotion experiences. Rather the evidence we provide refers to emotion mechanisms as they are represented in the shared meaning of emotion words used to reliably communicate emotion episodes and their characteristics to other people. As such, they represent very stable (rater reliability Alpha_C of .92 across 23 languages from most language families; Fontaine et al., 2013, p. 103) relationships for prototypical emotions. In contrast, self-report data on experienced emotion episodes vary very widely depending on individual appraisal patterns and reaction tendencies and being affected by multiple sources of error.

Importantly, the structure of effects we extracted from these data was found to be highly similar to the structure found in the study by Meuleman and Scherer (2013), which modelled real recalled emotion episodes. This lends support to the assumption of a substantial overlap between semantic and experience data, probably reflecting the effects of the lexical sedimentation mechanism proposed by Fontaine and colleagues (2013, Ch. 1) to explain such correspondence. Nevertheless, attempts should be made to replicate the present results for multicomponential data on actual emotion occurrences.

A potential limitation is that we did not explicitly model relations between the three response components (motivation, physiology, and expression). The set of variables of each component was reduced by PCA and then merely allowed to correlate by oblique rotations. While this was a reasonably flexible approach, we cannot exclude that vital nonlinear associations between the three response components were omitted by this procedure. Related to the PCA, we also must acknowledge that this procedure potentially introduced bias into the cross-validation for the MARS model. PCA was conducted before cross-validation, therefore different folds are not entirely independent, strictly speaking. However, we believe any bias that might have occurred this way to be minimal, since the

PCA was conducted as a pure dimension reduction step, and not tuned to give optimal results from the MARS analysis. Secondly, the loading structure uncovered by the authors of the original GRID study was found to be stable across different cultural samples and hence should be considered robust. We preferred basing our own analyses on this established structure, rather than re-establishing it by a cross-validated PCA.

A second limitation of the current study is the omission of *time* from our emotion model. Emotions are processes with temporal properties, such as onset, duration, and offset (Davidson, 1998). The operationalization that we have utilized in this study should be further modified by including time and the possibility of feedback relations. Such a modification is necessary if we wish to understand the full nature of emotion processes. We hope to address this issue in future studies.

A third limitation is that our analysis of clusters in component patterns—via simulation and nearest neighbor matching—remained at a largely qualitative and descriptive level. An objective quantification of goodness-of-fit was not possible due to the lack of ground truth labels for the simulated patterns. Nevertheless, the current simulations strongly suggest that our fitted model could be a powerful tool in generating concrete hypotheses and predictions to be validated in experimental research. At present, the testing of hypotheses about links between appraisals and other emotional components is still primarily based on theoretical prediction tables and decision trees. These predictions often concern only a limited set of appraisal criteria and emotion labels. A GRID-based simulation model could generate much more comprehensive and complex predictions about patterns of emotional responding than any theory could. Future studies could therefore benefit greatly from this type of computational hypothesis generation.

Conclusion. Our study represents the first attempt to examine emotion processes from a full-componential and interactional perspective using data on culturally shared semantic representations of the complex relationships between appraisals and different response components in emotion episodes. Both of these aspects constitute an important theoretical part of many appraisal theories but have received little empirical study. The present results show that accounting for interaction effects among appraisal criteria is essential for differentiating some—but not all—emotion patterns, and that a multicomponential approach provides a richer and more comprehensive characterization of what it means to be in an emotional episode. Apart from the presentation of a number of intrinsically interesting relationships, the findings reported here suggest several eminently testable hypothesis for studies examining actual emotion processes in everyday life as well as proposing some innovative methods for analysis and simulation.

Acknowledgment

This research was hosted by the University of Geneva and supported by the National Center of Competence in Research (NCCR) Affective sciences financed by the Swiss National Science Foundation (n° 51NF40-104897), Research Programme G.0223.13N of the Research Foundation - Flanders (FWO), and the Research Fund of KU Leuven (GOA/15/003).

5 References

Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, 74, 347–357.

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

Bates D., Maechler M., Bolker B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5.

Bossuyt, E., Moors, A., De Houwer, J. (2014a). Unexpected and just missed: The separate influence of the appraisals of expectancy and proximity on negative emotions. *Emotion*, 14 (2), 284–300.

Bossuyt, E., Moors, A., De Houwer, J. (2014b). On angry approach and fearful avoidance: The goal-dependent nature of emotional approach and avoidance tendencies. *Journal of Experimental Social Psychology*, 50, 118–124.

Brans, K., & Verduyn, P. (2014). Comparing determinants of emotion intensity and duration: Appraisals and regulation strategies. *PloS One*, 9, 1-13.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics.* Cambridge: Cambridge University Press.

Davidson, R. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion*, 12, 307–330.

Delplanque, S., Grandjean, D., Chrea, C., Coppin, G., Aymard, L., Cayeux, I., Margot, C., Velazco, M. I. S., Sander, D., & Scherer, K. R. (2009). Sequential unfolding of novelty and pleasantness appraisals of odors: evidence from facial electromyography and autonomic reactions. *Emotion*, 9(3), 316–328.

Ellsworth, P. C., & Smith, C. A. (1998). Shades of joy: Patterns of appraisal differentiating pleasant emotions. *Cognition and Emotion*, 2, 301–331.

Ellsworth, P., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–588). New York: Oxford University Press.

Fast, N. J., & Chen, S. (2009). When the boss feels inadequate: Power, incompetence, and aggression. *Psychological Science*, 20, 1406–1413.

Fitness, J., & Fletcher, G. J. O. (1993). Love, hate, anger, and jealousy in close relationships: A prototype and cognitive appraisal analysis. *Journal of Personality and Social Psychology*, *65(5)*, 942–958.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New York: Wiley

Fontaine J. R. J., Scherer K. R., & Soriano C. (2013). *Components of emotional meaning: A sourcebook*. Oxford University Press, Oxford.

Friedman, J. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19(1), 1–14.

Frijda. N. H. (1988). The laws of emotion. *American Psychologist*, 43. 349–358.

Frijda, N. H (1993). The place of appraisal in emotion. *Cognition and Emotion*, 7, 357–387.

Frijda, N. H. (2008). The psychologists' point of view. In M. Lewis, J. M. Haviland-Jones & L. Feldman Barrett (Eds.). *Handbook of emotions* (pp. 68–87). New York: Guilford.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, 57(2), 212–228.

Frijda, N. H., & Scherer, K. R. (2009). Emotion definitions (psychological perspectives). In D. Sander & K. R. Scherer (Eds.). *The oxford companion to emotion and the affective sciences* (pp. 142–144). New York: Oxford University Press.

Gentsch, K., Grandjean, D., & Scherer, K. R. (2013). Temporal dynamics of event-related potentials related to goal conduciveness and power appraisals. *Psychophysiology*, 50, 1010–1022.

Gentsch, K., Grandjean, D., & Scherer, K. R. (2014). Coherence explored between emotion components: Evidence from event-related potentials and facial electromyography. *Biological Psychology,* 98, 70–81.

Graham, T. R., Kowalski, K. C., & Crocker, P. R. (2002). The contributions of goal characteristics and causal attributions to emotional experience in youth sport participants. *Psychology of Sport and Exercise*, 3(4), 273–291.

Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, 8, 341–351.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Hosany, S. (2011). Appraisal determinants of tourist emotional responses. *Journal of Travel Research*, 51(3), 303–314.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. European Journal of Personality, 2, 171–203.

Kappas, A. (2001). A metaphor is a metaphor is a metaphor: Exorcising the homunculus from appraisal theory. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, and research* (pp. 157–172). Oxford, UK: Oxford University Press.

Kreibig, S. D., Gendolla, G. H., & Scherer, K. R. (2010). Psychophysiological effects of emotional responding to goal attainment. *Biological Psychology*, 84(3), 474–487.

Kreibig, S. D., Gendolla, G. H., & Scherer, K. R. (2012). Goal relevance and goal conduciveness appraisals lead to differential autonomic reactivity in emotional responding to performance feedback. *Biological Psychology*, 91(3), 365–375.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0-0. http://CRAN.R-project.org/package=lmerTest.

Lanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion*, 7(1), 207–212.

Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R. (2001). Relational meaning and discrete emotions. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 37–67). Oxford University Press.

Lewis, M. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, 28, 169–245.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. Scherer, T. Banziger, & E. Roesch (Eds.), *A blueprint for affective computing: A sourcebook* (pp. 21–46). US: Oxford University Press.

McGraw, K.M., (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology*, 53, 247–256.

Meuleman, B., & Scherer, K. R. (2013). Nonlinear appraisal modeling: An application of machine learning to the study of emotion production. *Affective Computing, IEEE Transactions on*, 4(4), 398–411.

Milborrow, S. (2011a). *earth: Multivariate Adaptive Regression Spline Models*. R package version 3.2-1. Derived from mda:mars by Hastie, T. and Tibshirani, R. http://CRAN.R-project.org/package=earth

Millborrow, S. (2011b). Notes on the Earth package.

Moors, A. (2009). Theories of emotion causation: A review. *Cognition and Emotion*, 23 (4), 625–662.

Moors, A. (2013). On the causal role of appraisal in emotion. *Emotion Review*, 5, 132–140.

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review*, 6, 303–307.

Moors, A., De Houwer, J. (2001). Automatic appraisal of motivational valence: Motivational affective priming and Simon effects. *Cognition and Emotion*, 15 (6), 749–766.

Moors, A., & Scherer, K. R. (2013). The role of appraisal in emotion. In M. Robinson, E. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 135–155). Guilford Press.

Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a dobby effect. *Emotion*, 9, 118–122.

Nguwi, Y., & Cho, S. (2010). Support-vector-based emergent self-organising approach for emotional understanding. *Connection Science*, 22(4), 355–371.

Nyer, P. U. (1997). A study of the relationships between cognitive appraisals and consumption emotions. *Journal of the Academy of Marketing Science*, 25(4), 296–304.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

Peters, A., & Hothorn, T. (2013). *ipred: Improved Predictors.* R package version 0.9-3. http://CRAN.R-project.org/package=ipred.

Reisenzein, R., & Spielhofer, C. (1994). Subjectively salient dimensions of emotional appraisal. *Motivation and Emotion*, 18(1), 31–77.

Revelle, W. (2013). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych Version = 1.3.2.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). Oxford University Press.

Roseman, I. J. (2013). Appraisal in the emotion system: Coherence in strategies for coping. *Emotion Review*, 5, 141–149.

Russel, D., & McAuley, E. (1986). Causal attributions, causal dimensions, and affective reactions to success and failure. *Journal of Personality and Social Psychology*, 50, 1174–1185.

Ruth, J. A., Brunel, F. F., & Ötnes, C. C. (2002). Linking thoughts to feelings: Investigating cognitive appraisals and consumption emotions in a mixed-emotions context. *Journal of the Academy of Marketing Science*, 29(1), 44–58.

Scherer, K. R. (1997). Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. *Cognition and Emotion*, 11, 113–150.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford University Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44 (4), 695–729.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23 (7), 1307–1351.

Scherer, K. R., & Ceschi, G. (1997). Lost luggage emotion: A field study of emotion antecedent appraisal. *Motivation and Emotion*, 21, 211–235.

Scherer, K. R., & Meuleman, B. (2013). Emotions can be predicted on theoretical grounds: Evidence from verbal labeling. *PLoS ONE*, 8 (3).

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2), 310–328.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Schliep, K., & Hechenbichler, K. (2014). *kknn: Weighted k-Nearest Neighbors*. R package version 1.2-5. http://CRAN.R-project.org/package=kknn

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation – different emotions : How appraisals shape our emotions. *Emotion*, 7(3), 592–600.

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

Sonnemans, J., & Frijda, N. H. (1995). The determinants of subjective emotional intensity. *Cognition and Emotion*, 9(5), 483–506.

Tong, E. M. W., Ellsworth, P. C., & Bishop, G. D. (2009). An S-shaped relationship between changes in appraisals and changes in emotions. *Emotion*, 9(6), 821–837.

Tong, E. M., Bishop, G. D., Enkelmann, H. C., Why, Y. P., Diong, S. M., Khader, M., & Ang, J. (2005). The use of ecological momentary assessment to test appraisal theories of emotion. *Emotion*, 5(4), 508–512.

van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and Emotion*, 18(5), 663–688.

Wei, T. (2013). *corrplot: Visualization of a correlation matrix*. R package version 0.73. http://CRAN.R-project.org/package=corrplot.