The GRID meets the wheel: assessing emotional feeling via self-report

Scherer, Klaus R.; Shuman, Vera; Fontaine, Johnny; Soriano, Cristina

How to cite

SCHERER, Klaus R. et al. The GRID meets the wheel: assessing emotional feeling via self-report. In: Components of emotional meaning: A sourcebook. J.J.R. Fontaine, K.R. Scherer & C. Soriano (Ed.). Oxford : Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199592746.003.0019

This publication URL: https://archive-ouverte.unige.ch/unige:97384
Publication DOI: 10.1093/acprof:oso/9780199592746.003.0019

Chapter 18

The GRID meets the Wheel: Assessing emotional feeling via self-report

Klaus R. Scherer,[1] Vera Shuman, Johnny J. R. Fontaine, and Cristina Soriano

The GRID study has provided a wealth of new data of high relevance to understand the semantics of emotion terms. This data can be profitably applied to create new tools for emotion research, or to further develop the existing ones. Here, we illustrate one such application by describing how the GRID paradigm was used to improve and further validate a popular tool for emotion assessment, namely the Geneva Emotion Wheel (GEW), a self-report measure of feelings.

Componential theories define emotion as a process during which several components such as physiological responses, motor expression, and cognitive representations (of both eliciting events and self-perceived response patterns) become synchronized over a limited period of time (see Chapter 1). One of these components is Subjective Feeling, a holistic cognitive representation that integrates the temporarily coordinated changes of the other components into a succinct, well formed Gestalt, allowing the individual to reach awareness of his/her state and label it—stating that he/she "has" or "feels" a particular emotion. In order to study the feeling component of emotion, psychologists need to rely on self-report. There is no other means but to ask the individual to report on the nature of his/her experience, since feeling is defined as a subjective cognitive representation of the emotional state which reflects a unique integration of mental and bodily changes in the context of a particular event (see Chapter 1). Emotion researchers currently use various paradigms for self-report, including the more recently developed GEW, to be reviewed below.

18.1 Classic self-report emotion assessment methods

Psychologists sometimes ask the participants in a study to describe their feelings in their own words. While this procedure may yield interesting information, it is fraught with problems. For example, people differ with respect to their verbal ability and richness of vocabulary (e.g., Gohm & Clore, 2000), which makes it difficult to compare reports across individuals or to rely on their accuracy. This is a major problem in controlled experimental research where fine-grained scalar measurement for a circumscribed number of specific feeling states is required. In consequence, psychologists generally use forced-choice self-reports of emotional experience. There are two major approaches: (1) the discrete emotion labels approach, and (2) the dimensional rating approach. The following section, based on Scherer (2005a), reviews these two traditions.

[1] Corresponding author: Klaus R. Scherer. Swiss Center for Affective Sciences—University of Geneva. 7, Rue des Battoirs, CH-1205 Geneva, Switzerland. Klaus.Scherer@unige.ch

The *discrete emotion labels approach* is used by scholars and laymen alike to categorize the stream of emotional experience into separate states profiting from the existence of specific emotion words and expressions in language (the type of words used in the GRID study). As shown in Chapter 3, while there are differences between languages with respect to the richness of the emotion lexicon and the meanings of related words, there is also a high degree of overlap. Darwin (1872) has used this convergence to postulate the evolutionary continuity of a set of fundamental emotions and the observable physiological and expressive symptoms that accompany them. This approach has been revived by Tomkins (1962), and has been popularized by Izard (1971; differential emotions theory) and Ekman (1972; basic emotion theory). In this tradition, categorizing emotional experiences according to the emotion words available in natural languages, it is typically assumed that the language-based categories reflect unique appraisal and response patterns (facial, vocal, and physiological) driven by typical event appraisals.

Researchers adopting the discrete emotion approach to assess emotional experience use scales with nominal, ordinal, or interval characteristics. Typically, respondents are presented with a list of emotion terms and are asked (1) to check the terms that best describe the emotion experienced (nominal scale); (2) to indicate on a scale (generally with 3 to 7 points) whether the emotion was experienced "a little," "somewhat," or "strongly" (ordinal scale); or (3) to use an analog scale (e.g., an underlying dimension from 0 to 100) to indicate exactly how much or how intensely the emotion was experienced (interval scale). Methods vary on whether respondents are to respond choosing only the most pertinent emotion category, two or more categories to indicate possible blends, or all categories in a list (replying with "no" or "0" for the emotions that are not at all appropriate to describe the experience). Some standardized instruments of this kind have been proposed in the literature (e.g., the Differential Emotion Scale; Izard, 1991). However, most researchers create ad hoc lists of emotion categories that seem pertinent for a specific research aim, without worrying too much about the representativeness of the chosen list or how well results obtained with the specific list may compare to results obtained with other emotion lists.

The results obtained with the emotion label approach are generally highly plausible and easily interpretable, given that widely shared and frequently used natural language labels tend to be employed. However, it is often difficult or even impossible to compare results across different studies in which widely different sets of emotion labels have been used. In addition, the statistical analysis of these data is hampered by the abundance of missing data and the difficulty to analyze and interpret the frequently reported emotion blends (Larsen et al., 2009; Scherer, 1998; Scherer & Ceschi, 2000). Often, problems are encountered of differential familiarity of respondents with particular emotion words, as well as differential interpretation of the meaning of the terms provided by the researcher. In addition, there are other problems with a discrete emotion response format, such as confusions (e.g., in the case of very extensive word lists), order effects, and other types of artifacts such as demand characteristics (e.g., the choice of specific emotion words may give away the research aim).

The *dimensional approach* in the self-report assessment of emotional experience was pioneered by the German psychologist Wilhelm Wundt (1896), who used introspection to develop a structural description of subjective feelings consisting of their position in a three-dimensional space formed by the dimensions of VALENCE (positive–negative), AROUSAL (calm–excited), and tension (tense–relaxed). This proposal has had a strong impact, both on the measurement of feeling (e.g., Schlosberg, 1954) and on the assessment of emotional connotations of language concepts in general (e.g., Osgood, Suci, & Tannenbaum, 1957). The theoretical foundations of this approach and its recent research development are discussed in detail in Chapter 2 of this book, showing that this domain of inquiry is currently dominated by a two-dimensional VALENCE × AROUSAL model.

To obtain a self-report of feeling with this approach, respondents are typically asked to indicate how positive or negative and how excited or aroused they feel (either in two separate steps, or by providing a two-dimensional surface and asking the respondent to determine the appropriate position in it; Larsen et al., 2009). As a result, the emotional feeling of the person is described by a point in this VALENCE × AROUSAL space. In some cases, respondents are also asked to separately evaluate the positive and negative parts of the VALENCE scale (see Chapter 2). In other cases, three dimensions (VALENCE, AROUSAL, and dominance) are assessed (e.g., Self Assessment Manikin Test, Bradley & Lang, 1994). This simple and reliable method lends itself to advanced statistical processing due to the general use of interval scaling. However, the information obtained is limited to the degree of positive or negative feeling and the degree of felt bodily excitation. Furthermore, the dimensional approach does not allow differentiating intensity of feeling from bodily excitation, which are clearly different constructs. For example, while intense anger is likely to be characterized by high AROUSAL, intense sadness is often characterized by very low AROUSAL. Thus, mild sadness and intense sadness could not be differentiated based on a VALENCE × AROUSAL space alone, as the AROUSAL level is low in both cases. This is a problem for researchers who are interested in clearly differentiating emotions like sadness and depression or grief. Obviously, any attempt to reduce positions in a multidimensional space to a two-dimensional representation will face this problem.

Another disadvantage is that, while most lay persons have little problem evaluating the positivity or negativity of a feeling (or event) and the approximate degree of their felt AROUSAL, the resulting point in two-dimensional space has no specific meaning for them and cannot be communicated to others. It would seem very strange to tell someone that I feel 2.3 positive and 1.6 aroused. Emotional sharing (Rimé, 2009) is an important social phenomenon, directly linked to the adaptive function of emotion communication through expression, and this function would not be well served by a two-dimensional metric.

Most importantly, however, in the two-dimensional VALENCE × AROUSAL space several rather different emotions are close neighbors. Figure 18.1 reproduces the mapping of the terms Russell (1983) used as markers for his emotion circumplex in two-dimensional VALENCE-AROUSAL space (original terms used by Russell in capital letters). In this figure, emotions as different as anger and fear, and other related terms, appear located in the immediate vicinity, as they all share high negative AROUSAL. In consequence, dimensional ratings on the VALENCE × AROUSAL space by themselves do not seem advisable if a researcher is interested in diagnosing the *quality* of a particular feeling. Reisenzein (1994) reviews this problem and suggests using appraisal theory to disambiguate the quality of neighboring states in VALENCE × AROUSAL space (Gehm & Scherer, 1988). As shown in this volume, the use of a discrete emotion label provides at least probabilistic information on the prototypical appraisal patterns differentiating the respective emotions. While appraisal profiles obviously provide the means for a more fine-grained differentiation (Scherer, Schorr, & Johnstone, 2001; and Chapters 1 and 12), this alone does not solve the problem of the lack of an appropriate self-report instrument that is precise, valid, and economical.

Clearly, the two classic self-report approaches reported above have both advantages and disadvantages. Given the central role of emotion self-report in emotion research, it is surprising that relatively few attempts have been made so far to develop new instruments to avoid some of their shortcomings. In particular, it might be worth investing in the development of an instrument capable of combining the advantages of the precise differentiation provided by natural language labels with the simple organizational structure afforded by a two-dimensional space. One possibility would be to use discrete emotion labels and arrange them graphically in a two-dimensional affect space, allowing the user to rapidly orient in this space. Such an instrument would be useful for a wide range of research goals.

Figure 18.1 Alternative dimensional structures for the semantic space of emotions.
Reproduced from Social Science Information, 44(4), What are emotions? And how can they be measured? Scherer, K. R., pp. 695–729, Figure 1 © 2005, Sage Publications, with permission.

18.2 **The Geneva Emotion Wheel—design features**

Based on these considerations, the Geneva Emotion Wheel (GEW) was designed to combine a discrete and a dimensional approach in the self-report assessment of emotion. It consists of a theory-based circular arrangement of discrete emotion terms in two-dimensional space according to the following design criteria (Scherer, 2005):

1 Two dimensions form the underlying structure of the instrument—VALENCE and control/POWER
2 The instrument uses standard emotion labels from natural languages to capitalize on respondents' intuitive understanding of the semantic field of emotion terms
3 Emotion terms are displayed in a systematic fashion by aligning them with respect to the underlying dimensional structure
4 The instrument allows the systematic assessment of the intensity of feelings
5 The instrument's user-friendly graphical design allows the respondent to rapidly understand the principle and use the instrument in a reliable fashion.

Below we describe how each of these design criteria was addressed in the development of the GEW:

1 **Choice of dimension.** There can be little doubt that VALENCE (positive–negative or pleasant–unpleasant), separating positive and negative emotions, constitutes the most important

dimension of affective space. The decision on the second dimension is less obvious. Despite the fact that, as shown above, the advocates of activation or AROUSAL as the second major dimension have dominated the scene for the last few decades, there are a number of drawbacks to this choice. As shown above, an AROUSAL dimension is very limited in its usefulness to differentiate between emotions due to the fact that most emotion families have several members that differ with respect to their degree of AROUSAL. Furthermore, it is not always clear exactly what is meant by the terms "AROUSAL" and "activation." While *AROUSAL* generally refers to sympathetic AROUSAL in the sense of the dominance of the sympathetic branch of the autonomous nervous system, *activation* is often used in connection with motivation or action tendencies, not necessarily requiring a high degree of sympathetic AROUSAL.

Beyond AROUSAL/activation, there are of course alternative dimensions that can be chosen as a second dimension after VALENCE. For example, Wundt (1905) proposed "tension–relaxation" and Schlosberg (1954) "attention–rejection." Most importantly, Osgood and collaborators (1957) highlighted the importance of a *potency* (or POWER, dominance) dimension in their seminal work on the semantic differential, placing this dimension second after *evaluation* (VALENCE), and before *activity* (activation, AROUSAL). This fits very well with an appraisal account of emotion. If emotions are indeed elicited and differentiated by appraisal patterns (see Chapter 1), the structure of the emotional space should be largely determined by the major appraisal dimensions. The close link between the appraisal checks "intrinsic (un)pleasantness" and "goal (in)consistency" or "goal conduciveness/obstruction" on the one hand, and the VALENCE dimension on the other, is obvious. The same is true for the link between the POWER/potency dimension and the coping potential check (which determines the degree of control available to the individual in a situation, as well as the POWER available to exercise that control; see Scherer, 1984b). As shown by numerous studies (see reviews in Ellsworth & Scherer, 2003; Scherer, Schorr, & Johnstone, 2001), the appraisal dimensions that seem to have the strongest impact on emotion differentiation are indeed goal conduciveness (representing VALENCE) and coping potential (control/POWER), corresponding to Lazarus' (1968) pioneering distinction between primary and secondary appraisal. Obviously, differences in AROUSAL/activation and intensity are also important determinants of feeling but they are more likely to define differences within an emotion family rather than between emotion families. In consequence, it was decided to use a two-dimensional affect space including the dimensions of VALENCE (based on pleasantness and goal conduciveness appraisals) and POWER/potency/control (based on coping potential appraisals) to organize the discrete emotion labels to be measured by the GEW.

Scherer (2005) further justified this theory-driven decision by arguing that this two-dimensional structure fits the organization of emotion terms in the two-dimensional space obtained, for example, through the analysis of similarity ratings. He provides an example, reproduced in Figure 18.1, in which a two-dimensional structure (conducive/obstructive × high/low control–POWER) found for 80 German terms (reported in Scherer, 1984b, pp. 47–55) is superimposed on the item distribution of English terms obtained by Russell (1983). The figure shows that both a classic VALENCE by AROUSAL structure (as postulated by Russell) and a VALENCE (conducive/obstructive) by control/POWER structure (as based on an appraisal model) can be justified, as the respective axes are just rotated by 45°. The latter structure provides a theoretically more homogeneous solution, as both factors represent the two major appraisal criteria.

2 **Choice of emotion families.** It is difficult to decide a priori, on purely theoretical grounds, which words, labels, or expressions should be chosen to represent the discrete states within the different regions of the continuous two-dimensional space spanned by VALENCE and POWER/potency. In the interest of replicability of results across studies, it was attempted to choose a

standard set of emotion families that met three conditions: (1) having been frequently used in past research (to ensure their utility in the assessment instrument), (2) covering most of the regions or segments of the two-dimensional space (to be able to map most positions in affective space), and (3) affording an arrangement of the terms around the rim of the wheel in approximately equal distances. As one might imagine, it is virtually impossible to find a set of emotion terms that equally satisfies all three conditions. In consequence, a pragmatic approach was adopted in the design of the GEW, giving greater weight to the first criterion to achieve a compromise that was satisfactory from the standpoint of potential users. Note that to accommodate users with specific needs, it is accepted that they replace part of the standard terms with terms of their choice, provided that the terms used are differentiated by VALENCE and control/POWER and can be reasonably represented on a circle in this space.

3 **Circular arrangement of emotion terms.** In the GEW, emotion terms representative of major emotion families are arranged in a circle (see Figure 18.3). The two underlying dimensions, VALENCE and control/POWER, separate the wheel into four quadrants: Negative VALENCE–low control/POWER, negative VALENCE–high control/POWER, positive VALENCE–low control/POWER, and positive VALENCE–high control/POWER. The position of the emotion terms in these clusters should correspond to their values on the VALENCE and control/POWER dimensions.

4 **Intensity ratings.** Members of any one specific emotion family can be expected to vary among each other with respect to intensity (e.g., irritation–anger–rage), which, as argued above, may correlate with, but is not the same as, physiological AROUSAL. It was, therefore, decided to map intensity on the distance between the rim and the hub of the wheel, representing the intensity of a specific emotional experience as the distance of its position from the central point in the VALENCE–control/POWER space (see also Reisenzein, 1994; Russell, 1980, p. 1170).

5 **Ease of use.** The wheel interface is easy to understand. Participants are asked to identify an experienced or imagined emotion among the various options provided. They are also asked to rate its intensity on the basis of the distance from the hub of the wheel, which implies choosing one of the answer circles increasing in size from the hub to the rim (the larger the circle, the more intense the emotion is reported to be). Thus, the meaning of the response options is quite intuitive. Also, in the interest of reading ease, the number of emotion families is limited. Finally, the alignment of the emotion terms based on the underlying dimensions should facilitate the usability of the GEW.

Note that the resulting instrument, although conceived in a very different fashion and for a rather different purpose, corresponds in several aspects to the various proposals of personality assessment instruments based on the notion of an "interpersonal circumplex", with the dimensions of "warmth vs hostility" or "love vs hate 'on the one hand and' dominance vs submission" on the other (Leary, 1957; Wiggins & Trobst, 1997). The two dimensions of nurturance/hostility (or warmth/coldness) and dominance/submission are highly comparable to VALENCE and control/POWER (which has often been called dominance in the literature). The arrangement of words in the GEW turns out to be also very similar to the emotion distribution in Plutchik's (1980b) emotion circumplex color wheel (see Figure 18.2), even though this theorist started from the notion of adaptation-oriented basic emotions. One might almost surmise that he arranged the emotions around the circumplex with an implicit VALENCE by POWER structure in mind.

In what follows, two stages of the development and of the investigation of the structural validity of the GEW will be described.

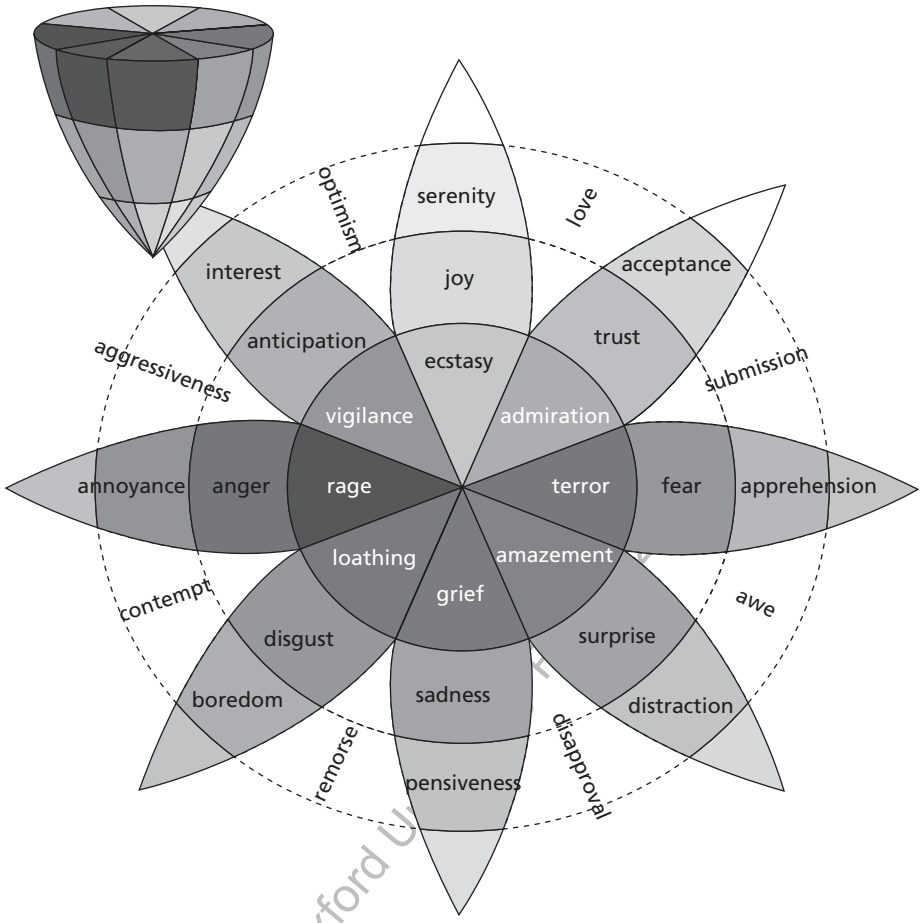

Figure 18.2 Plutchik's emotion circumplex.

Reproduced from American Scientist, 89, Plutchik, R., The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, pp. 344–350 © 2001, The Scientific Research Society. Reprinted by permission of American Scientist, magazine of Sigma Xi, The Scientific Research Society.

18.3 Stage 1: Development and structural validation of prototype versions of the GEW

The first prototype (Version 1.0) of the GEW (see Figure 18.3) was developed as a tool for the verbal report of emotions in a study of email communication (financed by the Gottlieb Daimler and Karl Benz Foundation). In this version, four emotion families were presented per quadrant, yielding a total of 16 emotions in the wheel (which seemed reasonable considering that the number of "basic emotions" is often considered to be somewhere between 6 and 14). In this version, a separate word (adjective) was proposed for each level of intensity response option within one emotion family (e.g., vexed, irritated, angry, enraged for the anger family; these are not visible in Figure 18.3, they appeared when the mouse cursor passed over the circles). The choice of the concrete families

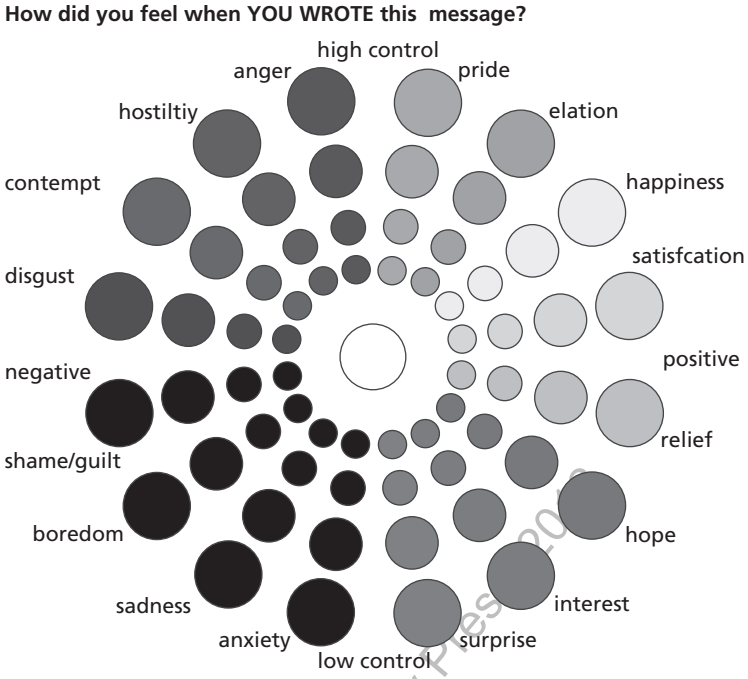

Figure 18.3 Version 1.0 of the Geneva Emotion Wheel. Reproduction of a VisualBasic screen dump; intermediate labels occurred when the mouse cursor passed over the circles.

Reproduced from Social Science Information, 44 (4), What are emotions? And how can they be measured? Scherer, K. R., pp. 695–729 © 2005, Sage Publications, with permission.

was largely determined by what are generally considered to be basic or fundamental emotions, frequently studied in the field.

Based on pilot studies with Version 1.0 of the GEW, a second version (Version 2.0, see Figure 18.4) was developed with two words (relatively close synonyms) referring to each of the 20 emotion families; the goal was to emphasize that each response option represented an emotion *family* rather than individual emotions. Furthermore, as the gradation of intensity levels by four different adjectives from the semantic field of the emotion family in the prototype Version 1.0 proved problematic (in terms of reliability and translatability of the gradation differences; Bänziger, Tran, & Scherer, 2005), in Version 2.0 the different intensity response options within one emotion family were therefore represented only with unlabeled circles of different sizes. Also, some emotions were placed in slightly different positions based on the results of the initial studies (e.g., "interest" passed to a somewhat higher position on the control/POWER dimension).

In several studies, the structural validity of placing the emotion terms in the GEW Versions 1.0 and 2.0 was assessed (see Bänziger, Tran, & Scherer, 2003, 2005; Sacharin, Schlegel, & Scherer, 2012). The results of these studies for the respective GEW Versions address similar issues and are, therefore, presented together below. Additionally, studies performed in our lab and other labs examined how well the GEW Versions 1.0 and 2.0 fared compared to other measurement tools (Caceido & van Beuzekom, 2006; Tran, 2004).

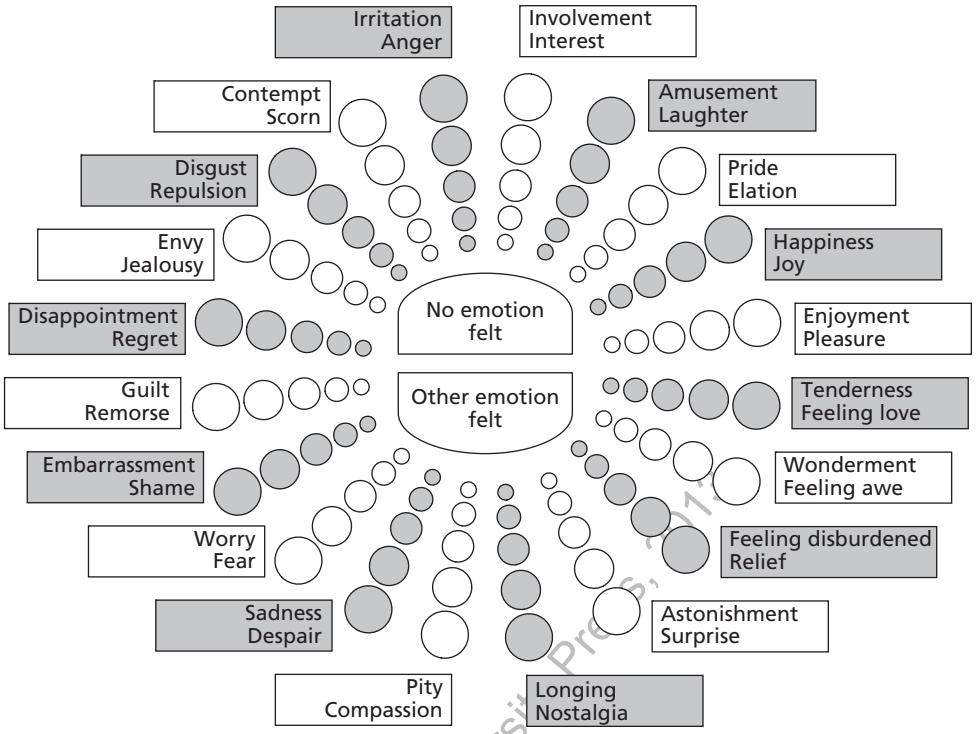

Figure 18.4 Template of Version 2.0 of the GEW (distributed until March 2013 via the website http://www.affective-sciences.org/researchmaterial and replaced by the new version 3.0 described in this chapter).

18.4 **Methods**

Two paradigms were used to assess the validity of the structure of the original version of the GEW: (1) similarity ratings of emotion words, and (2) direct ratings of the position of these words on the VALENCE and control/POWER dimensions. For Version 1.0, Bänziger et al. (2005) performed several validation tasks with a sample of 28 native English and 31 native French speakers. In a first task, respondents performed pair-wise similarity ratings of the 16 emotion family nouns positioned around the circumference of the wheel. In a second task, the 64 adjectives (16 × 4) representing the intensity gradations within families were sorted into the 16 family categories. In a third task, a subsample of 14 English and 15 French speakers rated the 16 emotion family nouns and 64 adjectives by using a dedicated graphic interface (enlarging a circle on the screen to rate intensity and moving markers within a two-dimensional space to rate VALENCE and control/POWER). Another subsample (14 English and 16 French speakers) rated the intensity, VALENCE, and control/POWER associated with each of the 16 emotion nouns and 64 adjectives on continuous rating scales using a mouse-operated slider. For Version 2.0, Sacharin et al. (2012) examined ratings of VALENCE and control/POWER on 11-point scales in an online study with 40 native English speakers. In this study, VALENCE was defined as follows: "the situation is experienced as (un)pleasant and enjoyable (disagreeable) and/or is likely to have positive and desired (negative and undesired) consequences for the person." In turn, control/POWER was described as "the person believes that he/she can (cannot) influence the situation to maintain or improve it (if desired)."

To examine the usability of the GEW, Tran (2004) compared the emotion assessment—first by using the GEW with ratings of emotions presented as word lists in a sample of 80 business school students, and later by administering the GEW in a simulation study with 106 managers attending executive development seminars in four to seven person teams. Caceido and van Beuzekom (2006) directly compared the utility of the GEW with the utility of the PrEmo (Desmet, 2003), a graphical measure for the assessment of discrete emotions.

18.5 **Results**

With regard to the structural validity of the GEW Version 1.0 (Bänziger et al., 2005), the categorization task showed that the adjectives representing intensity differences within an emotion family were almost all correctly classified. However, the agreement on the intensity gradation was less than perfect. With regard to the placement of the main emotion family labels in the VALENCE and POWER space, the ratings of similarity were submitted, for each language sample separately, to multidimensional scaling (MDS) analyses, using an ordinal model and Euclidian distances. The following fit measures for a two-dimensional solution were computed: English sample, stress = 0.34, RSQ = 0.52; French sample, stress = 0.34, RSQ = 0.39. The arrangement of the 16 emotion families in the two dimensions largely confirmed the theoretical prediction with respect to the overall clustering of the emotions in high/low POWER and high/low VALENCE. Exceptions were "interest" in the English sample and "relief" (soulagement) in the French sample, which were empirically found to be placed in the high control/POWER quadrant rather than the predicted low control/POWER quadrant. Similar to the MDS analyses, the data from the direct ratings of VALENCE and control/POWER supported the differentiated alignment of the emotions along the wheel very well for the VALENCE dimension. However, the empirical alignment of the emotions on the POWER/control dimension corresponded less well to the theoretical predictions. This was particularly true for the negative emotions. These emotions were also rated less reliably on the control/POWER dimension as indicated by large standard deviations.

The results of the Sacharin et al. (2012) study can be summarized as follows: for the VALENCE ratings, 19 out of 20 negative emotions and 15 out of 20 positive emotions were rated as predicted. Among the negative, only "compassion" was rated as more positive than expected. Among the positive, "nostalgia," "longing," "feeling disburdened," "astonishment," and "involvement" were not significantly different from the expected position. The results for the control/POWER ratings are more problematic—only 8 out of 20 high control/POWER emotions (all positive emotions) and only 1 out of 20 low control/POWER emotions ("sadness") were rated as predicted. After computing the mean VALENCE and control/POWER ratings for each word across raters, control/POWER and VALENCE were positively associated as reflected in a positive correlation of VALENCE and control/POWER ratings across the 40 word ratings, $r(40) = 0.718$, $p < 0.001$. Furthermore, control/POWER ratings had greater standard deviations than VALENCE ratings (2.89, 1.51), $t(39) = 11.68$, $p < 0.001$. This is in line with the findings reported by Bänziger et al. (2005) with respect to the difficulty of finding the predicted arrangement for the negative, low control/POWER emotions.

Studies examining the usability of the GEW for emotion assessment compared with other measures showed that the same feelings are associated with different vignette scenarios when using the nouns in GEW Version 1.0 as word lists (Tran, 2004). For the use with managers in a quasi-naturalistic environment, very high response rates were obtained in daily assessments for 8 to 10 days, providing an indicator of the managers' strong involvement in the emotion assessment task. The GEW was judged to be a particularly useful measurement instrument under

time pressure and with repeated measurements (Tran, 2004). Furthermore, in Caceido and van Beuzekom's (2006) study, respondents overall preferred the GEW over the PrEmo, and judged the GEW as clear to understand, useful to differentiate between emotions, and appealing in its visual design.

18.6 Discussion

The studies examining the structural validity of the emotion terms in the VALENCE–control/POWER space showed that for GEW Versions 1.0 and 2.0 the placement of the emotion terms along the VALENCE dimension generally corresponded to prediction. To improve the representation, "compassion," which had been rated as a positive emotion, was to be moved from the negative to the positive side of the GEW.

In contrast to the findings for the VALENCE dimension, the alignment of emotion terms on the control/POWER dimension differs depending on the response paradigm. In the similarity study with MDS analyses, though the model fit was not very good and the sample size was small, the placement of the emotion terms corresponded well to the predicted alignment. In contrast, the empirical data derived from rating studies tend to deviate from the predictions, especially in the negative—low control/POWER quadrant. Inspection of means and variance suggest that, to some extent, the rating results could reflect a response bias in the use of the control/POWER dimension by some participants, resulting in a large variance (Bänziger et al., 2005; Sacharin et al., 2012).

Using a larger sample size or re-wording the description of what control/POWER means might ameliorate this problem. The discussion of the results for the appraisal component in the GRID study (see Chapter 12) also showed that the wording of the control/POWER features in the questionnaire may not have been optimal, and changes are proposed for the CoreGRID and further studies with the full instrument (see Chapter 44). However, it remains to be seen if the changes in wording produce the desired effect. It may well be that the abstract notions of "control" and "POWER" in connection to emotions are not easy to grasp for non-psychologists, and that it is thus difficult to obtain explicit and reliable ratings for these concepts.

An alternative explanation for the observed difficulty in obtaining the theoretically predicted alignment on the POWER dimension (especially for negative high-power and positive low-power emotions) is provided by the strong association of VALENCE with control/POWER ratings (Sacharin et al., 2012). Indeed, it has recently been suggested that control/POWER appraisals are valenced (Shuman, Sander, & Scherer, 2013; Scherer, 2010b). High POWER is associated with positive affect and low POWER with negative affect (e.g., Keltner, Gruenfeld, & Anderson, 2003). A similar correlation was found in the results of the GRID study, as shown in Chapters 8 and 12, suggesting a strong ecological correlation between negative VALENCE and low control/POWER.

Thus, even with a larger sample size and revised wording, it may not be possible to obtain independent and fine-grained ratings of emotions on the control/POWER appraisal criterion alone. To empirically grasp the notion that negative and positive emotional experiences can be associated with more or less control/POWER, other methods may be needed that measure not only control and POWER, in general, or specific appraisals related to it, but additional components of the emotional experience, such as action tendencies.

A further substantiation of the structure of the GEW is timely because the GEW is increasingly used due to its user-friendliness in comparison with other instruments (Caceido & van Beuzekom, 2006). The GEW has been applied in a variety of contexts, such as consumer attitudes to

internet videos and industrial design products (e.g., Bardzell, Bardzell, & Pace, 2009; Caceido & van Beuzekom, 2006), the affective evaluation of body movements and vocalizations (e.g., Beck, Stevens, & Bard, 2009; Pammi & Schröder, 2009), emotions during learning in virtual environments and in virtual environments with different illumination (e.g., Longhi, Pereira, Brecht, & Behar, 2009; Santos, 2008), and experience sampling studies of emotions in everyday life (e.g., Tschan, Semmer, Messerli, & Janevski, 2010).

Furthermore, the GEW has been applied to assess emotions at different levels of analysis ranging from the individual and the team level emotions of managers (Tran, 2004) to the emotional climate in a hospital (Wittgenstein, 2008). Specifically, Tran (2004) found that the emotions recorded in the different quadrants of the GEW differentially influence team decision making and cohesion. For example, negative–low control/POWER emotions were positively associated with team cohesion, and negative–high control/POWER emotions were negatively associated with team cohesion. Furthermore, differentiating between the different intensity levels within an emotion family in the GEW contributed important information. For example, moderate levels of positive–low control/POWER emotions were positively associated with alternative evaluation (a key component of decision making), whereas high intensity positive–low control/POWER emotions were negatively associated with alternative evaluation. Finally, Tran's work showed that the GEW can be used as a means to help develop team processes. Over the course of her study, it was observed that participants often used the GEW ratings as a basis to discuss their emotions with their colleagues, yielding self-awareness and awareness at the group level. In addition, by mapping their emotions on the GEW on a regular basis, all group members can see the evolution of the emotional climate and can proactively manage it (Tran, Páez, & Sánchez, 2012).

To conclude, the existing studies on the GEW Versions 1.0 and 2.0 underscore the utility of the instrument. The placement of the emotion terms in each quadrant of the GEW, however, could not be satisfactorily justified to date. MDS analyses of similarity ratings indicate that the placement of the emotions predicted in each GEW quadrant are valid, but direct rating studies of the VALENCE and control/POWER of the same terms were unable to yield the expected placement on the control/POWER dimension. Ratings of control/POWER, it seems, have to be worded in drastically different ways to obtain results that reflect the MDS results.

18.7 Stage 2: Development of a standard version of the GEW based on structural validation with the GRID instrument

After the first results of the GRID study had confirmed that the two major dimensions of the semantic emotion domain are indeed VALENCE and control/POWER (Fontaine et al., 2007), it became obvious that the GRID paradigm could constitute a royal road to finalize the validation of the GEW for a set of major emotion terms. As demonstrated in the chapters of Part III of this volume, the GRID results clearly establish the existence of a four factor structure for the emotion space, with activation/AROUSAL and NOVELTY being essential additional factors for a satisfactory mapping of major emotion terms with respect to their discriminability in low-dimensional space. However, we decided to stick to a two-dimensional representation of the emotion terms in a wheel structure for the assessment instrument, as a three-dimensional representation on a two-dimensional paper or screen surface is confusing, and a four-dimensional representation would require a series of independent two-dimensional graphs. Such formats are inacceptable for a self-report instrument that needs to be immediately obvious to use and economical in terms of time investment. As the two first dimensions emerging in PCAs of the GRID data, VALENCE and POWER also explain the largest percentage of the variance between emotions, this solution seems well justified.

The GRID study described in this volume provides an ideal framework to obtain similarity metrics for emotion words. The GRID provides a very comprehensive feature profile consisting of 142 features covering all components of emotion. In consequence, the assessment of similarity is based on a very rich set of criteria. Furthermore, the information is obtained for a very large set of languages and cultures using sizable groups of native speakers. In consequence, all requirements to obtain a definitive validation of the placement of the emotion terms around the circumference of the wheel are fulfilled. Most importantly, the POWER dimension emerges as the second strongest dimension from the PCA and is clearly identified even if the appraisal component is not included in the analysis. This consistent with the original choice of control/POWER as the second factor in the two-dimensional structure of the GEW (in contrast to AROUSAL in dimensional theories). It is also consistent with the assumption that the appraisals of control and POWER strongly affect the other components producing clear changes, such as dominant action tendencies and loud voice that are sufficient to determine a clear, overall POWER factor. In consequence, the use of VALENCE and POWER coordinates derived from the GRID data seem to be an ideal solution to solve the issue of validating the predicted arrangement of the GEW emotion terms in two-dimensional VALENCE by control/POWER space.

To this end, it was necessary to obtain additional ratings for words used in the GEW that were not rated in the basic GRID study. The list of 24 GRID emotion words (see Chapter 5 for the criteria of choice) already contained 16 of the 20 items that had been regularly used with the previous versions of the GEW. Four words were missing: "amusement," "admiration," "relief," and "regret." Therefore we contacted the different collaborators in the GRID study and asked them whether they would be willing and able to have these words judged on the 142 features in the same way and using similar groups of participants as for the regular GRID questionnaire. Groups in ten different countries—Switzerland (French), United Kingdom (English), Belgium (Flemish), China (Mandarin Chinese), Germany (German), Estonia (Estonian), Finland (Finnish), Italy (Italian), Japan (Japanese), Poland (Polish)—agreed to participate and collected the data using exactly the same procedures as described in Chapters 5 and 6 (see Table 18.1 for sample characteristics).

Table 18.1 Sample characteristics

Language	Country	Region	N			Age			Format	Compensation
			Total	F	M	Range	Mean	SD		
French	Switzerland	Geneva	20	12	2	20–45	29.21	8.285	online	course credit
English	UK	Belfast & York	19	10	9	18–22	19.32	1.416	online	course credit
Dutch	Belgium	Gent	15	9	6	21–23	21.87	0.640	online	course credit
Chinese	China	Hong Kong	19	6	11	19–25	21.24	1.602	online	course credit
German	Germany	Berlin	20	19	1	19–37	22.10	4.241	online	course credit
Estonian	Estonia	Tartu	15	8	7	22–29	25.67	1.759	online	course credit
Finnish	Finland	various	18	17	1	18–37	25.33	5.087	online	course credit
Italian	Italy	Bologna	19	12	2	19–26	23.36	1.946	online	course credit
Japanese	Japan	Sapporo	15	9	6	18–21	18.87	1.125	online	course credit
Polish	Poland	Lodz	14	7	7	19–38	23.36	4.749	online	course credit

In consequence, the new standard version of the GEW presented here contains the following 20 emotion words (the asterisks indicating the four words rated specifically for the GEW validation):

- admiration*
- amusement*
- anger
- compassion
- contempt
- contentment
- disappointment
- disgust
- fear
- guilt
- hate
- interest
- joy
- love
- pleasure
- pride
- regret*
- relief*
- sadness
- shame

We first analyzed the reliability of the rating data for the four new terms for each country (following the procedure described in Chapter 6). These data were then combined with the data for the 16 GEW terms that were part of the basic GRID study and for which the reliability had been assessed previously. Then, dissimilarity matrices were produced for each of the country samples by computing the distances between the feature profiles of the different words. These matrices were then combined and submitted to MDS (using Proxscal) across countries. Proxscal computes solutions for different dimensionalities. As we are interested in a two-dimensional arrangement of the emotion terms for the GEW instrument, here we report only the results for the two-dimensional solution. The stress and fit measures for the two-dimensional solution are as follows: Normalized Raw Stress 0.00481, Stress-I 0.06939, Stress-II 0.13588, S-Stress 0.00503, Dispersion Accounted For (D.A.F.) 0.99519, Tucker's Coefficient of Congruence 0.99759. These fit indices can be judged as quite satisfactory in the light of the levels expected according to the MDS literature. Table 18.2 shows the respective coordinates for the 20 words on the two dimensions (shown as a two-dimensional plot in Figure 18.5).

The two dimensions underlying the plotted coordinates of the 20 GEW words can be readily interpreted—the horizontal dimension corresponding to a general VALENCE dimension (highly comparable to the VALENCE dimension in the overall GRID analysis—see Chapter 7), separating the group of positive emotions from that of the negative ones. The vertical dimension is not as immediately obvious, but comes very close to the POWER dimension in the overall GRID analysis (see Chapter 7). Both of these dimensions correspond directly to the dimensions that have been theoretically

Table 18.2 Coordinates of the 20 GEW emotion words in two-dimensional VALENCE by POWER space

Emotion words	VALENCE	POWER
admiration	0.66	–0.09
amusement	0.67	0.19
anger	–0.37	0.47
compassion	–0.05	–0.55
contempt	–0.55	0.43
contentment	0.77	–0.03
disappointment	–0.77	–0.12
disgust	–0.68	0.20
fear	–0.61	0.07
guilt	–0.57	–0.27
hate	–0.45	0.43
interest	0.61	0.25
joy	0.68	0.07
love	0.58	–0.16
pleasure	0.71	0.02
pride	0.72	0.15
regret	–0.70	–0.19
relief	0.66	–0.36
sadness	–0.68	–0.35
shame	–0.61	–0.16

postulated for the GEW. They seem to be closely related to two types of appraisal: on the one hand, appraisals of unpleasantness and obstructiveness vs pleasantness and conduciveness (underlying the VALENCE dimension), and on the other hand, appraisals of high vs low control/POWER (for the POWER dimension) (see Chapter 12). We also computed a PCA of these data to compare the MDS plots to factor score plots. Essentially, the PCA extracted four factors, the first two of which again correspond to VALENCE and POWER. The two-dimensional factor score plot yielded coordinates for the 20 emotion words that were very similar to the MDS solution. A comparison of the positions of the 20 emotion words in the two-dimensional MDS space shows that the patterning is very comparable to the placement of the same words in the previous Version 2.0 of the GEW shown in Figure 18.4, thus empirically confirming the earlier arrangement, which was based exclusively on theoretical considerations.

Based on these results, a new version (3.0) of the GEW was constructed. Concretely, the two-dimensional coordinates shown in Table 18.2 and Figure 18.5 were projected onto the rim of the circle that represents the wheel structure and arranged equidistantly around the circumference of the wheel. The result of this empirical GRID-based circular ordering of the 20 terms, GEW Version 3.0, is shown in Figure 18.6.

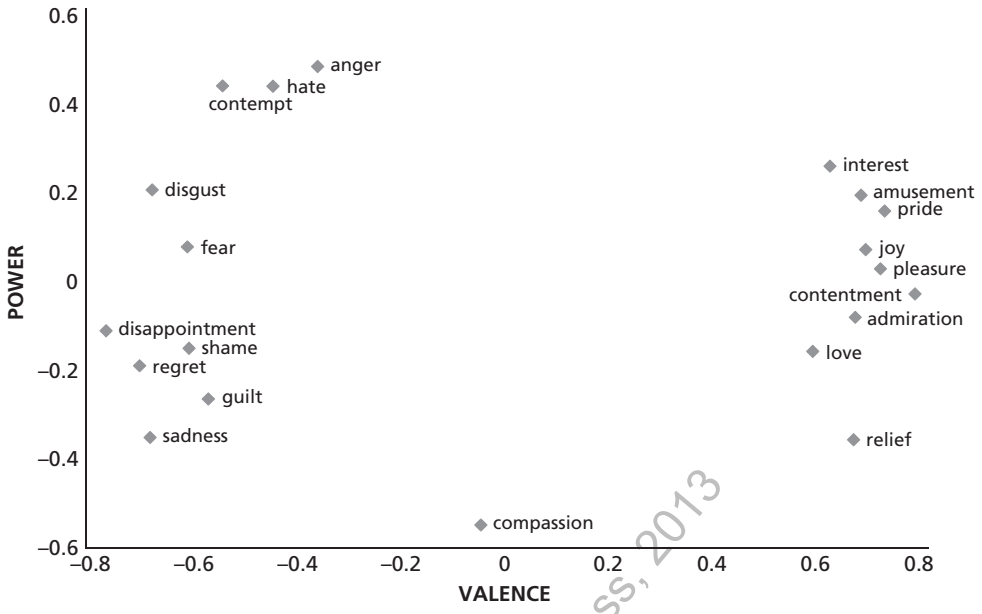

Figure 18.5 Representation of the 20 GEW emotion words in a two-dimensional VALENCE by POWER space.

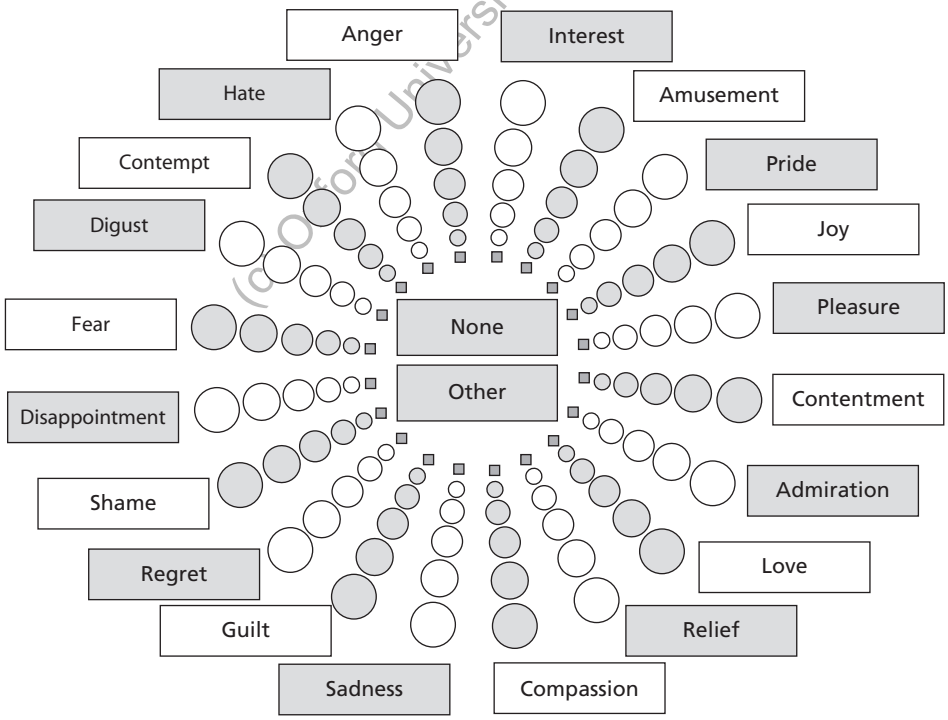

Figure 18.6 Template of Version 3.0 of the GEW as based on the GRID validation described in this chapter.

18.8 Availability and analysis procedures

The most recent version of the GEW (Version 3.0) shown in Figure 18.6 is now available for non-commercial use by academic researchers interested in self-report assessment of emotion. The recommended instructions, the final version of the wheel interface, and further information (about adaptation and translation into other languages and licensing for commercial use of the instrument) are available on the GEW web page (http://www.affective-sciences.org/gew). There, the wheel can be downloaded in Word format that can be easily adapted to the needs of different researchers. The web page also shows information about the currently available versions for different languages and computer applications of the wheel.

The 20 items currently used in the wheel, like the 24 items in the GRID study, have been chosen on the basis of an extensive selection process based on theoretical and empirical considerations. They represent a fair sampling of the more frequently used emotion terms in different languages (see Chapter 5 in this volume for the GRID choices). However, they may not constitute an optimal choice for *all* possible applications in different fields of research. For example, an event sampling of emotions occurring in families may require a different selection of emotion categories than an experimental study with a limited set of emotion manipulations. Given the huge diversity of research interests and needs in the field, it would be illusory to propose the use of the current standard version for all kinds of application.

It should be noted that the specific choice of emotion words is not constitutive for the GEW and the advantages its use confers to researchers. Rather, the fact that the arrangement of the terms in two-dimensional space is theory-based and empirically confirmed makes the instrument much easier to use than the usual lists (especially in repeated applications, as participants will find the appropriate terms much more rapidly and precisely). The use of the instrument is also facilitated by the anchoring of the meaning of the chosen emotion terms with respect to their position in the underlying, theoretically determined, affective space. In consequence, users are invited to construct their own wheel if the choice of terms in the standard version is not optimal for the respective research aim (see instructions for adaptation on the GEW web page provided above). However, we would like to stress that, whenever possible, it is advantageous to use the standard version shown in Figure 18.6 to allow for replication by other researchers and to build up cumulative databases.

Finally, we provide a brief overview of the analysis procedures. The GEW can be analyzed in two different ways:

1 Using the classic discrete emotion approach, the ratings on different emotion families in the GEW structure can be analyzed in a very similar fashion to the procedures used with standard questionnaires in which emotion terms are listed one below the other and participants are asked to rate the intensity with which they have experienced each of the emotions on a five-point ordinal scale. Depending on the purpose of the study, participants can be asked to (a) select only one emotion family (the strongest they experienced), (b) choose several emotions they may have experienced simultaneously or in close succession (producing mixed emotions), or (c) give a response to each emotion scale, with a special category for "did NOT experience this emotion" (a format that has desirable psychometric characteristics).

2 Using a dimensional approach allows the researcher to obtain continuous dimensional values by profiting from the explicit arrangement of the emotion terms in two-dimensional VALENCE by control/POWER space. Concretely, each emotion term can be represented by its two coordinates in the space (provided in Table 18.2) with the respective intensity as a third variable. These

three values can then be analyzed separately or in the form of a composite scale. In the case of mixed emotions, with several emotions rated, the respective coordinates and intensities can be combined using statistical measures of central tendency. Alternatively, all ratings within one or more of the four quadrants of the wheel can be combined with the help of statistical measures of central tendency.

18.9 **Conclusion**

The GEW is a theory-based instrument for the assessment of emotional experience through self-report. It combines a dimensional orientation with ratings of intensity for a number of major emotion families, rated in a categorical fashion. The instrument has been used for many years by emotion researchers in many different applications. Earlier studies with multidimensional analyses of similarity ratings of emotion words (Bänziger et al., 2005), designed to validate the theoretical structure that underlies the arrangement of the emotion families on a wheel-like circle, have essentially confirmed the predicted placement of the emotion families in a two-dimensional space formed by VALENCE and control/POWER. However, attempts to use direct ratings of the appraisals that are thought to determine the differentiation of emotions on the control/POWER dimension have met with mixed success. Two of those studies (Bänziger et al., 2005; Sacharin et al., 2012) have been reviewed and it is suggested that it might be useful to use a different approach to validate the position of emotion words in a VALENCE × POWER structure, abandoning the exclusive reliance on coping potential appraisals in favor of a more comprehensive representation of POWER.

The GRID paradigm has been found to be ideally suited for this purpose, given that it anchors the four fundamental dimensions of emotional space in feature profiles based on all emotion components, and given the remarkable stability of these factors over many different languages. The four dimensions reliably emerge even when individual components (also appraisal) are removed from the data set (see data reports in Part III of this volume). The analysis of the 20 GEW words with the GRID instrument in 10 countries with 10 different languages has allowed us to firmly validate the theoretical structure used in the development of the wheel. The precise data provided by these GRID results have allowed recalibrating the positioning of the emotion categories on the rim of the emotion wheel. Based on this recalibration, a new version of the GEW, 3.0, is presented in this chapter and is made available to interested researchers. While the theoretical structure of the GEW has been validated in this research, an empirical construct validation remains to be done.

Apart from providing a combination of a dimensional and a categorical approach to emotion assessment through self-report, the GEW provides an alternative to the dominant VALENCE × AROUSAL model in dimensional approaches. We hope that the new Version 3.0 of the GEW will see widespread use in the future and will help standardize the self-report assessment procedures in the psychological investigation of emotion and in the interdisciplinary domain of the affective sciences more generally. We believe that systematic use of a standard version of the GEW would constitute important progress with respect to replicating results across studies, and, most importantly, allow increased collaboration in designing empirical studies and sharing the data in the emotion domain. These activities, in turn, would greatly contribute to our effort to better understand the semantics of emotion terms and their role in the categorization and communication of emotional experience.