BEYOND VERBAL SCALES: MEASUREMENT OF EMOTIONS IN ADVERTISING EFFECTIVENESS RESEARCH
by Oliver Hupp, Andrea Gröppel-Klein, Anja Dieckmann, Philipp Broeckelmann and Kathrin Walter*

ABSTRACT
Emotions are a major topic in marketing science and practice alike (Kroeber-Riel/Weinberg/Gröppel-Klein 2008, p. 99). B2B, B2C, and retail brands need to be emotionally charged to ensure differentiation from competitor brands. But how can emotions be reliably measured? Only valid and meaningful measurement of emotions will allow the emotional impact of marketing activities to be optimized in the future.

The Institute for Consumer and Behavioural Research at Saarland University together with brand and methodology specialists at GfK joined forces to develop a picture-based scale that is applicable at various brand touch points: it is called GfK EMO Sensor. This article summarizes the theoretical background and the development of the methodological approach. It concludes by presenting initial applications in advertising effectiveness research and their practical implications for marketing.

* Dr. Oliver Hupp, Division Manager Brand and Communication Research, GfK AG, Nordwestring 101, D - 90319 Nürnberg; email: oliver.hupp@gfk.com
Prof. Dr. Andrea Gröppel-Klein, Professor of Marketing and Director of the Institute for Consumer and Behavioural Research, Saarland University, Campus A5 4, D - 66123 Saarbrücken; email: groeppel-klein@ikv.uni-saarland.de
Dr. Anja Dieckmann, Senior Specialist in the Department of Basic Research, GfK-Nürnberg e.V., Nordwestring 101, D - 90319 Nürnberg, E-Mail: anja.dieckmann@gfk.com
Dipl.-Kfm. Philipp Broeckelmann, Research Assistant at the Institute for Consumer and Behavioural Research, Saarland University, Campus A5 4, D - 66123 Saarbrücken; email: broeckelmann@ikv.uni-saarland.de
Dipl.-Kffr. Kathrin Walter, Junior Specialist Advanced Statistics & Methodological Consulting, GfK AG, Nordwestring 101, D - 90319 Nürnberg; email: kathrin.walter@gfk.com

The authors would like to thank Prof. Dr. Frank Wimmer for his helpful comments during the review process, and also Daniel Becker and Holger Dietrich for their valuable feedback on earlier versions of this article. We are also very grateful to Dr. Dorothea Baun and Dr. Raimund Wildner for their important input during the development process of GfK EMO Sensor.

1. Introduction

Emotional experiences shape people’s attitudes towards brands and products. The pleasure of indulging in one’s favorite chocolate, or the pride experienced when friends admire one’s new sports car strengthen our brand commitment. Yet it is dampened by frustration at spending endless time in the waiting loop of a service hotline. Thus, emotions continuously affect brand differentiation and long-term brand commitment (cf. Keller 2008a). Among the manifold brand touch points, emotions elicited by communication and, in particular, advertising are especially important for marketing practice (cf. Haimerl 2007). In his seminal work “Visual Communication”, Kroeber-Riel (1993, p. 155 ff.) presents numerous examples of both highly successful and weak emotional stimuli in advertising, and explains their effects on brand perception and brand attitude.

In the GfK Advertising Effectiveness Model, which is based on work by Keller (2008a, 2008b), emotions elicited by the advertising draw the audience’s attention to the advertising media and thus, ultimately, affect the success of the advertising (for a simplified illustration, see Figure 1).

In the model, catalysts determine the extent to which advertising can lead to changes in the customer mindset (1). These catalysts serve as filters, and can strengthen or weaken the impact of advertising. An advertising stimulus needs to pass through these filters and – in an initial, partly unconscious processing step – be classified as personally relevant in order to gain sufficient attention to be taken on to further processing (cf. Fransen 1999, p. 41 ff.; Heath 2007, p. 9). In the Anglo-American marketing lite-
rature, this process is called engagement, while in Germany the term involvement is more common (2). The impact of advertising on brand perception differs depending on engagement strength. The weakest level of advertising impact is the anchorage of a certain advertisement and the advertised brand in memory. Stronger impacts, in ascending order, are effects on specific brand associations (brand identity, brand personality, brand benefits), and also on the judgment of the advertised brand, and finally there is the impact on brand loyalty and purchase intention.

As mentioned above, emotional stimuli can act as catalysts to intensify engagement and thus influence the effects of advertising on the customer’s mindset. First, stimuli that produce emotional reactions will prompt the audience to orient itself towards the triggering event (“primary attention”). Afterwards, strong emotional stimuli can induce the audience to pay greater attention to the advertisement for a longer time (“secondary, sustained attention”). Thus, emotions are crucial factors in contributing to success, affecting whether an advertising stimulus will gain recognition in the competitive environment and be further processed.

The monitoring of emotional reactions to advertising therefore plays a major role in advertising effectiveness research: What emotions are elicited by the people, music and scenes presented in advertisements? And how strong is the emotional reaction? Answering these questions is highly important to not only measure, but also to explain changes in attitude and behavior – a central goal of advertising effectiveness research. Gaining insights into consumers’ emotional reactions will thus facilitate a targeted and effective management of brand touch points.

2. Theory of emotions

2.1 What are emotions?

Emotions are of the utmost importance in marketing. But what exactly is an emotion? “Everyone knows what an emotion is, until asked to give a definition”. This quotation from Fehr and Russell (1984, p. 464) illustrates the difficulty science and marketing have in capturing the content of the construct emotion, and distinguishing it from closely related constructs. This difficulty contributes to a considerable degree of uncertainty of how emotions can be validly measured.

Until now, there has been no uniform, broadly accepted definition of emotions. Rather, most current definitions are confined to a list of characteristics (cf. Scherer 2005; Plutchik 1991) according to which, emotions are complex reactions to events that are relevant to a person’s subjective wellbeing. They involve changes in physiology, behavior (including facial expressions), and subjective experience. Emotions depend on the
conscious or unconscious appraisal of the triggering event and are associated with certain behavioral responses, which are partly formed by evolution. They can be distinguished from other affective phenomena by their relatively fast onset, short duration and high intensity (3).

Basically, one can distinguish two different traditions in emotion theories: **Appraisal theorists** propose that emotions result from the comparison between a desirable or undesirable goal and the degree of goal attainment. Lazarus (e.g., 1982), one of the most important appraisal theorists, considers cognitive activity a necessary precondition for emotions. According to his theory, emotional experience requires the understanding that an event can have positive or negative consequences for subjective well-being. Also, people can anticipate and revive emotions by pure mental activity, which in turn can lead to physical reactions.

According to **biologically oriented emotion theorists**, emotions are genetically programmed (e.g., Zajonc 1980). They can be triggered automatically and without higher cognitive processing by innate or learned stimuli, even before one becomes conscious of these stimuli (e.g., LeDoux 1996). With the help of modern brain-imaging technologies, neuroscientists have in recent years done impressive work in shedding light upon such automatic emotional processes. A series of studies now demonstrates that emotions can bias decisions in a certain direction before any conscious thought process occurs (cf. Bechara et al. 2005; Damasio 1996). Arguments which have been traditionally regarded as rational are often presented as mere post-hoc rationalizations (e.g., Haidt 2001). Impulse buying, for instance, can be associated with emotions such as happiness and surprise (cf. Baun/Gröppel-Klein 2003).

However, based on neurological studies, LeDoux (1996) demonstrates that emotions can also result purely from mental anticipation and construction. Most current emotion researchers thus regard both theoretical accounts as legitimate. Emotions can occur automatically and unconsciously, facilitating fast reactions (e.g., “freezing”). But they can also result from conscious assessments in cortical areas. There, it is assessed whether relevant goals have been attained or not and whether they are associated with rewards or punishments (for more detail, see Kroger-Risd et al. 2008, p. 118 f.). But if emotions can be both biologically pre-wired and arise from higher cognitive activity, their measurement in marketing and advertising effectiveness research becomes even more challenging.

### 2.2 Primary and secondary emotions

In a large part of the emotion literature so-called basic emotions are distinguished from more complex secondary emotions. Going back to Darwin (1872), based on his observations of emotional expressions in many different human societies and across different...
species, he postulated the existence of a number of fundamental, innate emotions. Most scientists today still assume a set of primary or basic emotions (e.g., Ekman 1972; Tomkins 1962; Plutchik 1991; Izard 1971). They are characterized by special qualities, such as cross-cultural transferability, developed in earliest childhood, and certain physiological, mimic and behavioral reactions that inextricably accompany them. Some scientists propose that more complex secondary emotions are the result of the blending of basic emotions (cf. Izard 1994; Plutchik 1962, 2003).

Measurement of emotions

How can emotions be validly measured? Can they simply be captured by self-report or standardized verbal scales? Biologically oriented emotion researchers often apply neuroscientific methods to measure physiological indicators of emotions such as heart rate, skin conductance level, or regional brain responses (cf. Kroeber-Riel et al. 2008, p. 120 ff.). Observational methods for capturing emotional facial expressions are also popular, such as the Facial Action Coding System (FACS; cf. Ekman/Friesen 1975). These approaches have obvious advantages. They are highly objective and can capture emotions while they are occurring. But there are also disadvantages, especially when it comes to their application in marketing. For instance, not all emotions have a unique physiological or mimic pattern. Especially secondary emotions can hardly be deduced in an unambiguous way. Additionally, application of these methods requires much time, effort and high levels of expertise, making them almost impracticable for large samples.

Appraisal theorists mainly concentrate on capturing the conscious emotional experience by verbal methods, such as questionnaires or think-aloud techniques. Indeed, self-report provides the only access to the subjective experience level of emotions, despite all the progress in the neurosciences. The variety of more complex secondary emotions cannot be revealed in any other way. Additionally, questionnaires are easy to apply and cost-effective, both in data collection and analysis. But they can lead to problems as well. Verbal methods are often criticized for inducing rationalizations in respondents and preventing intuitive responses. This raises serious concerns: Do respondents have sufficient insight into their emotions to be able to answer questions truthfully? Does the interview situation induce cognitive filtering of emotional experiences, thereby increasing the tendency to give socially desirable answers? Which unconscious or automatic components of the emotional reactions are missed by verbal measurement approaches?

Picture-based approaches, unlike purely verbal questionnaires, are expected to facilitate intuitive access to emotions that are difficult to verbalize or which respondents are
reluctant to talk about. Krober-Riel et al. (2008, p. 124) recommend picture scales especially for the following applications:

- “for modality-specific measurement of pictorially encoded emotions”,
- to gain “access to feelings that are difficult to verbalize or not explicitly conscious”,
- when “the risk of socially desirable answers is to be diminished”, as well as
- in “low-involvement situations” (e.g., when viewing TV ads).

At the same time, picture-based scales can give a fresh impetus to the interview and increase interest, motivation, and, ultimately, response rates. For the basic emotions, picture scales have the additional advantage of cross-cultural applicability. Finally, ease and efficiency of application are comparable to verbal scales.

However, selection and validation of the pictures for such scales requires special care. It is very difficult to find truly unambiguous pictures that represent the same emotion for most people, especially if the scale is to be extended beyond the basic emotions. For instance, misinterpretations can result when the depicted emotional scene is not taken as metaphor for an emotion (e.g., the picture of a broadly-smiling woman on a merry-go-round is not seen to represent “joy”), but is only associated with the concrete situation (e.g., “being in an amusement park”). Therefore, it needs to be examined whether verbally labeling pictures or collages of pictures with an appropriate short verbal description might increase the clarity and power of the message. Additionally, idiosyncratic associations with the pictures can lead to distorted answers. A sleeping cat might express comfortable warm feelings for some people, while others will only think of their allergies. Finally, pictures can themselves trigger emotions. In psychological research, for instance, an internationally validated set of pictures (IAPS: Internationaal Affectieve Picture System; cf. Lang/Bradley/Cuthbert 1999) is often used to induce emotions. Pictures therefore pose the risk of changing the emotional experience to be measured. When selecting pictures, it is therefore important to make as sure as possible that the picture only shows an emotion (e.g., a frightened woman), but does not induce it (e.g., close-up view of a hairy spider).

In summary, there is no gold standard yet with which all aspects of emotions can be captured (cf. Scherer 2005, p. 709). It is therefore recommended that pragmatic criteria should be taken into account when deciding in favor of a certain measurement approach. Our goal is the development of a measurement tool which validly captures consumers’ manifold emotions, regardless of whether they are basic or more complex secondary emotions. Also, it needs to be widely applicable in cross-cultural and online studies, as well as for different products and at diverse brand touch points.
3. **GfK EMO Sensor: Development of a scale for capturing emotional responses**

The number of emotions whose marketing relevance has been demonstrated in numerous studies extends far beyond the basic emotions (for an overview, see Bagozzi/Gopinath/Nyer 1999; Kroeber-Riel et al. 2008, p. 132 ff.). As explained above, only self-report provides access to the differentiated world of human emotional experiences. Thus, researchers at GfK in cooperation with the Institute for Consumer and Behavioural Research at Saarland University decided to develop a questionnaire-based methodology, flexible in its application, for capturing discrete emotions at various brand touch points.

As mentioned above, both verbal and picture-based methodologies have pros and cons. The final decision for one approach – verbal or pictorial – will thus be based on an empirical comparison. To allow for a fair comparison, special attention needs to be paid

---

**Figure 2**

<table>
<thead>
<tr>
<th>Step-by-step development of GfK EMO Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1:</strong> Selection of relevant emotions</td>
</tr>
<tr>
<td>• Selection of basic emotions as per Plutchik</td>
</tr>
<tr>
<td>• Supplemented by emotions important in a marketing context</td>
</tr>
<tr>
<td><strong>Step 2:</strong> Selection of pictures to represent the emotions</td>
</tr>
<tr>
<td>• Picture library search</td>
</tr>
<tr>
<td><strong>Step 3:</strong> Validation of the collected picture material</td>
</tr>
<tr>
<td>• Validation study A: Unaided categorization of sets of pictures and construction of ranking</td>
</tr>
<tr>
<td>• Validation study B: Cross-validation of pictures using film-clips</td>
</tr>
<tr>
<td>• Validation study C: Unaided sorting of pictures followed by ranking</td>
</tr>
<tr>
<td>• Validation study D: Cross-validation by measurement of electrodermal reaction</td>
</tr>
<tr>
<td><strong>Step 4:</strong> Comparison of different scale formats</td>
</tr>
<tr>
<td>• Recording of emotional reactions to film ads in an online survey using various scale formats</td>
</tr>
</tbody>
</table>
to the selection of appropriate, unambiguous pictures. The development of the scale for capturing emotions can therefore be divided into different steps (see Figure 2). First, relevant emotions are selected to be included in the scale. Second, a search is mounted for pictures to represent each of these emotions. In a third step, the picture material is reduced in a series of validation studies to only the most unambiguous pictures. Finally, different scale variants – verbal, pictorial, and combinations of them – are used to capture emotional responses to TV ads and then evaluated on different performance criteria.

### 3.1 Step 1: Selection of relevant emotions

Choosing from the large number of emotions mentioned in the relevant literature is not a trivial task. (The fundamental PAD scale alone comprises more than 150 items describing emotional states, cf. Russell/Mehrabian 1977.) To derive a set of emotions that included as many emotions as necessary but at the same time as few as possible, the concept of basic emotions devised by Plutchik (2003) was used as a starting point. Plutchik reduces all emotions to eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) with different degrees of intensity, and mixtures of them. Love, for instance, is seen as mixture of the basic emotions of joy and trust. The eight basic emotions were supplemented by eight “mixed” emotions (primary dyads: aggression, awe, disappointment, love, optimism, remorse, submission, contempt) as well as by emotions that have been shown to be especially relevant in a marketing context. These were boredom, compassion, coziness, desire, embarrassment, envy, interest, longing, passion, pride, rejection, and superiority (e.g., Aaker/Stayman/Hagerty 1986; Belk/Ger/Askegaard 2003; Bosch/Schiel/Winder 2006; Verbeek/Belschak/Bagozzi 2004; Zeitlin/Westwood 1986). In the end, the list consisted of 28 emotions to be tested.

### 3.2 Step 2: Selection of pictures to represent the emotions

With emotion words as search terms, more than 10,000 pictures from commercial picture libraries were inspected and evaluated in terms of their suitability for representing the selected emotions. Special attention was paid to the criterion that a picture should not induce the relevant emotion in the respondent, but only depict it. The emotional experience of the viewer of the pictures should thus remain as unaffected as possible (cf. Bosch et al. 2006). For each emotion, a set of 10 to 15 pictures was selected, which was later reduced to between 5 and 7 pictures per emotion based on expert judgment; they aimed for balance in the depicted situations, as well as in the gender, age and ethnic background of the depicted persons (see Figure 3 for an example). This resulted in a set of 152 pictures in total.
3.3 Step 3: Validation of the collected picture material

To assess the suitability of the set of pictures and to further reduce its size, the pictures underwent a series of validation studies.

3.3.1 Validation study A

The 28 collages of 5 to 7 pictures per emotion were presented consecutively to a sample of 80 participants. They were asked to provide an unaided labeling of the 28 collages. Content analysis revealed good to excellent identification rates ($\geq 70\%$ to $\geq 90\%$ agreement among respondents in terms of the depicted emotion). Only the collages for the emotions awe, optimism and longing showed a need for improvement ($\geq 60\%$ agreement). Based on these results, some pictures in these categories were replaced by new pictures.
Forty of the participants were next asked to rank the pictures in each collage in terms of how well each represented the emotion they saw in the collage. For 26 out of the 28 collages, a clear rank order could be determined, suggesting comparable perception of the pictures by the participants.

### 3.3.2 Validation study B

Validation was not restricted to verbal methods, but also included validation based on visual material to maintain modality specificity (cf. Kroeber-Riel et al. 2008). Specifically, film clips (without sound) with clear-cut emotional content were selected to represent the eight basic emotions. For the emotion of anger, for instance, a scene from the movie “Pulp Fiction” was chosen (see Figure 4).

The selected film clips were shown to 283 participants. Subsequently, participants were asked to assign one picture out of a set of different emotional pictures to each film clip, and then give a verbal description of the emotion they had seen in it. The assignment of pictures to film clips was in line with the researchers’ expectations by a large majority of participants ($\geq 82\%$) in seven of the eight emotions. Only for the basic emotion of anticipation, was the proportion of participants lower who agreed that this was indeed the emotion depicted in the movie scene (61%). Results of the verbal de-

---

**Figure 4**

**Bruce Willis in “Pulp Fiction”**
scription of the movie also showed satisfying agreement levels, but the spectrum of answers was generally broader, supporting the notion that pictures might be more suitable for measuring emotional experiences.

### 3.3.3 Validation study C

In validation study A, the assignment of pictures to emotions had followed the researchers’ judgment, and pictures were presented together in collages accordingly. Next, the aim of validation study C was to find out whether respondents would sort the pictures in accordance with their expected emotional content even without previous knowledge of the emotional categories. To address questions of this kind, Griffin and Hauser (1993) developed a methodology that does not provide respondents with pre-selected categories, but captures the dimensions and relationships that respondents themselves see in the presented items. The task is to sort stimuli according to their perceived similarity into as many piles as participants find reasonable. A second goal of the study was to further reduce the picture sets to three pictures per emotion.

A representative sample of 501 participants (51% female; M age = 42.6 years, SD age = 14.6) was invited to take part in personal interviews at GfK test studios in Berlin, Dresden, Essen, Frankfurt, and Munich. To avoid cognitive overload, each participant was only shown pictures of half (i.e., 14) of the emotion categories, selected on a random basis. Participants were told that the pictures represented emotions. Their task was to sort pictures into piles so that in their view, the pictures in each pile represented the same or almost the same emotion. The participants were not given any further guidelines in terms of number of piles or number of pictures per pile. After completing the sorting, they were asked to choose the most representative picture from each pile, and then to label each pile – first unaided, then by allocating little tags that had the relevant emotions pre-printed on them.

For each pair of pictures, the percentage of participants who had sorted the two pictures into the same pile ("co-assignment rate"). These percentages were used as proximity measure in a hierarchical cluster analysis. This analysis resulted in clear clusters of pictures for 25 of the 28 expected emotion categories (Ward’s procedure). This means that 25 clusters consisted precisely of those pictures that were also included in the pre-defined emotion category. Three of these 25 emotion clusters were excluded because the mean paired co-assignment rates were below 50% (optimism, remorse, contempt). The pictures intended to represent the emotions awe, longing and submission ended up in different clusters and thus, no coherent cluster was derived for these emotions. Across the remaining 22 emotions the mean paired co-assignment rate was 71%. Based on analysis of answers in the unaided labeling task, the emotion category
anticipation was relabeled curiosity. All other categories received the intended or a closely related label from the majority of participants.

The selection of the pictures was based on two central criteria for each picture:
- number of votes for being the most representative picture of an emotion within a correctly labeled picture pile as a measure of power to express an emotion
- assignment rate to a pile labeled with the intended emotion as measure of its unambiguousness

In terms of the first criterion, the three pictures with most votes for being most representative were selected. The second criterion of unambiguousness was only used to ensure that the most representative pictures were not ambivalent. However, the two criteria only rarely produced diverging rank orders, thus corroborating the quality of the selection. Moreover, the selected pictures were consistent with those chosen as best pictures in validation study A in 77% of the cases, so the solution can be regarded as stable.

Table 1

<table>
<thead>
<tr>
<th>Emotion cluster</th>
<th>av. paired co-assignment rate</th>
<th>Emotion cluster</th>
<th>av. paired co-assignment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggression</td>
<td>68%</td>
<td>Fear</td>
<td>82%</td>
</tr>
<tr>
<td>Anger</td>
<td>89%</td>
<td>Interest</td>
<td>64%</td>
</tr>
<tr>
<td>Boredom</td>
<td>79%</td>
<td>Joy</td>
<td>85%</td>
</tr>
<tr>
<td>Compassion</td>
<td>88%</td>
<td>Love</td>
<td>88%</td>
</tr>
<tr>
<td>Coziness</td>
<td>82%</td>
<td>Passion</td>
<td>74%</td>
</tr>
<tr>
<td>Curiosity (formerly anticipation)</td>
<td>63%</td>
<td>Pride</td>
<td>82%</td>
</tr>
<tr>
<td>Desire</td>
<td>71%</td>
<td>Rejection</td>
<td>88%</td>
</tr>
<tr>
<td>Disgust</td>
<td>89%</td>
<td>Sadness</td>
<td>95%</td>
</tr>
<tr>
<td>Disappointment</td>
<td>71%</td>
<td>Superiority</td>
<td>59%</td>
</tr>
<tr>
<td>Embarrassment</td>
<td>71%</td>
<td>Surprise</td>
<td>73%</td>
</tr>
<tr>
<td>Envy</td>
<td>72%</td>
<td>Trust</td>
<td>57%</td>
</tr>
</tbody>
</table>
In summary, the results of the sorting task strongly suggest a reduction of the theoretically derived emotion categories to 22 (see Table 1), each represented by three pictures, selected in terms of power to express an emotion and unambiguousness. The reduction of the emotion clusters to only the three selected pictures increases the mean paired co-assignment rates from 71% to 77%, averaged across all 22 emotions.

### 3.3.4 Validation study D

A final validation study examined the relationship between the evaluation of TV ads using the emotional pictures on the one hand, and electrodermal reaction during viewing on the other. Electrodermal activity is an indicator of strong emotional arousal (cf. Boucsein 1992; Gröppel-Klein 2004). One group of 41 participants were shown a high-arousal TV ad for DeBeukelaer Prinzenrolle biscuits, in which the Prince tries to kiss Sleeping Beauty awake. Another group of 39 participants were shown a less arousing TV spot for the same product, in which four cooks prepare the chocolate cream for the biscuits. Out of a pile of 84 pictures (the three best pictures indicated by validation study C, for all 28 originally selected emotions), participants selected those pictures that best represented their emotions while viewing the spots. The levels of the electrodermal reaction for the spots had been established in an earlier study by Gröppel-Klein/Domke/Bartmann (2006).

As expected, participants assigned more emotion pictures to the Sleeping Beauty spot (7.6 pictures on average) than to the cooks spot (6.8 pictures); in particular, they assigned more positive emotion pictures to the Sleeping Beauty spot (4.3 positive emotion pictures on average, compared to 3.3 for the cooks spot). Besides supporting the general validity of the pictorial measurement approach, the study revealed some difficulties concerning the selection of pictures by participants. Pictures that showed similar situational motifs to those used in the TV ad were assigned to the spot based on their outward similarity, regardless of their emotional content. For instance, two of the pictures representing the emotion rejection were not once assigned to the Sleeping Beauty spot (see Figure 5, left and center). One picture, however, that showed a kneeling man holding flowers in his hands who is rebuffed by a woman (see Figure 5, right) was assigned to the spot by 34.1% of participants. Obviously, although Sleeping Beauty in the TV spot is enraptured by her prince and does not reject him, this particular rejection picture was selected by a considerable number of participants due to its situational similarity to the TV spot. Just like in the picture, the spot shows roses and a man courting a woman. Thus, in future studies it has to be ensured that pictures are not used that show situations similar to motifs used in the advertising campaign that is to be evaluated.
Step 4: Comparison of different scale formats

Based on the studies already completed, a validated set of emotion pictures is now available. But in what format should they be administered to reveal respondents’ emotional reactions? In a first practical test, different presentation formats were used to capture emotional responses to TV ads, and compared with each other on different performance criteria. Based on expert judgment, five TV spots were selected that were expected to elicit very different emotions:

- An image spot for Adidas, with highly aggressive content, depicting a soccer theme,
- a commercial for Axe deodorants, with disgusting but at the same time funny content,
- a spot for Landliebe yoghurts, depicting happy family life in the countryside,
- a matter-of-fact commercial for HUK-Coburg car insurances, dominated by the price issue, and finally
- a spot for Havana Club rum, showing young people partying on the streets while Latin music is playing.

3.4.1 Design

The main goal was to compare five different formats (see Figure 6) for presenting the 22 emotions. A primary aim was to compare verbal items and pictures:

- Format 1: purely verbal emotion labels (“words”)
- Format 2: purely pictorial representation of emotions by individual pictures (“pictures”)
This comparison was to reveal whether the above-mentioned advantages of pictures (such as access to emotions that are difficult to verbalize) outweigh their disadvantages (such as a certain degree of ambivalence that remains despite the thorough validation process).

Three additional variants represented different attempts to foster unambiguous interpretations of the picture material:

- Format 3: Purely pictorial representation of emotions by collages consisting of the three single pictures selected for each emotion ("collages"). The presentation of three pictures was expected to reduce ambivalence by revealing the intended emotion as the common theme that unites the pictures.

- Format 4: Combining verbal and pictorial stimuli in labeled single pictures, that is, pictures that are subtitled by the appropriate emotion word ("pictures + words")
Format 5: Combining verbal and pictorial stimuli in labeled collages, that is, collages that are subtitled by the appropriate emotion word ("collages + words")

By comparing labeled picture variants to purely pictorial variants on the one hand, it could be found out if the addition of verbal labels indeed increases unambiguousness. On the other hand, the comparison of labeled picture formats and purely verbal formats would reveal if the presentation of vivid, intuitive pictures in addition to the words yields any advantages in terms of the performance criteria. Participants were randomly assigned to one of the five format conditions.

To further facilitate intuitive answers, a simple dichotomous answer format was used. Participants were asked to indicate for each item (i.e., for each emotion) whether they had felt the emotion during the TV ad, yes or no.

The following performance criteria were used to evaluate the different scale variants:

- Scale handling (evaluation of understandability, user friendliness, and enjoying using the scale, as seen by the interviewees themselves)
- Validity in terms of conformity with expert judgments
- Differentiation between emotions
- Differentiation between TV ads

### 3.4.2 Sample

The sample consisted of 1250 participants (50% female; \( M_{\text{age}} = 40.4 \) years, \( SD_{\text{age}} = 14.4 \)) from AskGfK, the GfK online panel. For her co-operation, each respondent received 250 panel points worth 2.50 Euros.

### 3.4.3 Procedure

At the beginning of the interview, respondents were presented with a list of brands from each of the five product categories for which TV ads were shown. They were asked to select those brands that they would consider buying and to indicate their first choice. Subsequently, each respondent saw two ads. Immediately after each ad, participants completed the emotion scale, presented in the format that had been randomly assigned to them. After completing the emotion scale, some general questions about the ad were asked. After both TV ads had been shown and evaluated, participants again indicated the brands they would consider buying as well as their first choice for each product category. At the end, respondents were asked to evaluate how understandable they had found the emotion scale, and how much they had enjoyed doing the task.
3.4.4. Results

- Scale handling – Evaluation of understandability, user friendliness, and enjoying using the scale

The unlabeled picture and collage formats were considered to be almost equal in terms of scale handling, with one exception: With collages, it was less likely that none of the items could express the emotion that participants actually felt during the spot (see Figure 7). The purely verbal word format was evaluated as superior to the unlabeled formats in terms of ability to assign and express the emotions felt during

<table>
<thead>
<tr>
<th>Evaluation of scale handling by participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Using the ... I could assign my feelings very easily</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I found it very simple during the TV ad to express my feelings in this way</td>
</tr>
<tr>
<td>None of the provided ... actually expressed what I felt during the ad (inverted)</td>
</tr>
<tr>
<td>Unfortunately I was not able to fully understand the ... (inverted)</td>
</tr>
<tr>
<td>I found the survey very enjoyable and interesting</td>
</tr>
<tr>
<td>I found the questions about the feelings created by the TV ad very boring (inverted)</td>
</tr>
<tr>
<td>I enjoyed the survey</td>
</tr>
<tr>
<td>Scale handling (mean value of all statements)</td>
</tr>
</tbody>
</table>

* p < 0.10 / ** p < 0.05

Figure 7
the ad. Using the scale was further improved by presenting pictures in addition to verbal labels, that is, by the word plus picture format, especially regarding enjoyment, interest in the task, and ability to assign emotions.

Using labeled collages instead of labeled single pictures did not improve scale handling any further. Participants even perceived the labeled collage format as slightly more tiring, less interesting than labeled single pictures. This result confirmed the comparison between unlabeled single pictures versus collages, which did not reveal clear added benefits for collages.

Significant differences are illustrated only for neighboring scale formats in Figure 7. Further significant contrasts exist, especially between labeled and unlabeled variants.

Validity – agreement with expert judgments

To evaluate the validity of the emotion profiles resulting from the five different scale formats, 18 brand and communications experts at GfK indicated the emotions they expected to be elicited by each spot. Figure 8 shows mean absolute deviations, averaged across emotions, and correlations between the mean emotion profiles of respondents and those of the GfK experts. Again, the unlabeled formats were outperformed by the labeled formats, both with and without pictures. The word plus picture format achieved the highest level of agreement with the experts’ judgments.

As already seen with the Sleeping Beauty spot in the picture validation study D, unlabeled formats raise further validity concerns if picture motifs bear a resemblance to motifs shown in the TV ads. Picture assignment is then sometimes based on the outward similarity in appearance of motifs (e.g., of the depicted persons) instead of being based on the conformity of the emotions elicited by the ad with those expressed by the picture. For instance, the emotion disappointment was assigned to the Adidas spot much more frequently in the unlabeled than in the labeled picture con-

<table>
<thead>
<tr>
<th>Agreement of respondents’ emotion profiles with expert judgments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictures</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Mean absolute deviation from expert judgments</td>
</tr>
<tr>
<td>Correlation with expert judgments</td>
</tr>
</tbody>
</table>
A more detailed analysis based on the assignment frequencies for each of the three pictures shows that the frequent mention of disappointment can be ascribed to the fact that a sports scene is shown on two of the disappointment pictures. The unlabeled picture of a disappointed track athlete was assigned by 35% of respondents, that of a disappointed goal-keeper by as many as 42%, while only 8% selected the picture of a disappointed man on the telephone. In contrast, in the labeled picture conditions assignment rates dropped to 13%, 7%, and 6%, respectively. Similar problems with unlabeled pictures could be observed for basically every ad that was tested. The provision of labels seems to facilitate recollection of and adherence to the instruction to assign pictures based on the depicted emotion.

- **Differentiation between emotions**

A frequent criticism of standard verbal rating scales is their lack of differentiation between different items and between different evaluated objects (such as brands). Picture-based scales are expected to produce more pronounced differentiation. Ideally, to derive clear recommendations for practitioners, certain emotions should be clearly assigned while others should be clearly dismissed. To evaluate the differentiation potential of the different scale formats, it is useful to compare the standard deviations of the assignment rates across the different emotions (see Figure 9).

Assignment rates in the labeled conditions have higher overall standard deviations compared to unlabeled conditions and thus provide a more clear-cut differentiation between emotions. For a more detailed analysis, the 22 emotions were divided into two groups according to their valence: positive emotions (coziness, curiosity, desire, interest, joy, love, passion, pride, surprise, trust) and negative emotions

<table>
<thead>
<tr>
<th>Ability of the five scale formats to differentiate between emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average standard deviation across</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>… all emotions</td>
</tr>
<tr>
<td>… positive emotions</td>
</tr>
<tr>
<td>… negative emotions</td>
</tr>
</tbody>
</table>
For the positive emotions in particular, the difference between labeled and unlabeled conditions was even larger. The relatively low standard deviations in the unlabeled conditions could be interpreted as another hint that the emotional content of certain pictures is not unambiguously recognized, resulting in random responses and thus indeterminate assignment rates rather than in clear yes or no trends. Focusing on the negative emotions only, it is the purely verbal format that has the lowest standard deviation of all formats. In the light of this result, the mean assignment levels for negative emotions also reveal an interesting, potentially unwelcome trend: The purely verbal format achieves the lowest assignment rate of 9.8% on average across all negative emotions, whereas all other formats achieve average assignment rates in double figures (15.0% on average). Taken together, the low assignment rates and the poor differentiation of the word format for negative emotions suggest a stronger trend towards socially desirable answers. In purely verbal questionnaires, participants may be more reluctant to admit negative emotions.

Figure 10

**Positioning of spots in an emotional map, based on the five different scale formats**
Differentiation between TV ads

To evaluate the ability of the different scale formats to differentiate between spots, *multidimensional scaling* was applied. This way, the position of each spot could be plotted onto an *emotion map* separately for each format condition (see Figure 10). The further apart a spot is located from the center, the more different it is from the other spots in terms of elicited emotions. Figure 10 shows that the word format leads to the least pronounced differentiation: The spots are closer together in the emotional map when their emotional impact has been measured using words only, compared to all other formats. For the spots located in the positive region on the left side of the emotion map, the labeled variants (labeled pictures and labeled collages) lead to especially pronounced differentiation. For the more negative right side of the emotion map, all pictorial formats, labeled and unlabeled, perform comparably well in producing a differentiated positioning for the spots. This result is in line with the above-mentioned ability of the different formats to differentiate between emotions.

Summary

The scale formats with labels (words only, labeled pictures, labeled collages) outperform the purely pictorial formats (pictures, collages) on almost all comparison criteria. An especially serious problem is that unlabeled pictures lead to the assignment of some emotions that are completely unexpected in terms of the expert judgments. This can be explained by similarities in appearance between picture motifs and motifs shown in the TV ad. This problem can be avoided by using labeled pictures.

The comparison between the two labeled pictorial formats and the purely verbal format reveals several advantages from presenting a picture in addition to a verbal label. Labeled pictures lead to more pronounced differentiation across both emotions and TV spots. Socially desirable answers can be reduced, as indicated by higher selection rates for the negative emotions that a spot is expected to elicit in terms of the expert judgments. In general, labeled pictures produce higher levels of agreement with expert judgments and the scale handling gets better scores compared to the other formats.

Regarding the comparison between labeled single pictures and labeled collages no conclusive recommendation can be derived. By providing a broader spectrum of pictures, labeled collages can reduce the risk that pictures are selected due to similarity in appearance between the emotion picture and the object to be evaluated. But on the other hand, the large number of pictures can make the interview more boring for interviewees and lead to declining attention. It would be advisable to assess this risk in pre-tests, or alternatively to move to using the labeled single picture from the outset.
In sum, pictures that are subtitled with the appropriate emotion label form an ideal compromise for capturing emotions. Thus, as far as the central question is concerned of whether verbal or pictorial methods are superior, neither one nor the other is superior; instead, a combination of words and pictures outperforms alternative formats across a wide range of performance criteria. Misinterpretations and wrong selections based on outward similarities of depicted motifs can thus be avoided. At the same time, the provision of pictures obviously increases the willingness of respondents to admit to experiencing negative emotions. This reduces the detrimental effects of social desirability, encourages honest and spontaneous answers, and thus produces differentiated profiles even for negative emotions. The scale – named GfK EMO Sensor – can now be applied to a wide range of questions relevant in marketing.

Applications

GfK EMO Sensor can help to reveal what lies beneath consumer attitudes and behavior. With this research tool, marketers can explore the emotions elicited at different brand touch points such as ads, exhibition stands, services, and products, and receive answers to questions of great practical relevance: What kind of emotions are elicited by certain marketing activities, and what is their valence? How intense is the emotional experience induced in the audience? Does the marketing achieve truly integrated communications by yielding coherent emotional effects using different marketing activities?

For advertising agencies, valuable recommendations can be derived from target-group-specific analysis of GfK EMO Sensor data. For instance, the Axe TV ad that was included in the large comparison study described above induced mainly negative emotions – especially rejection and disgust – in viewers over 30 years old. In the product’s target group of young men, however, the ad succeeded in inducing also positive emotions, in particular joy and interest, probably due to its humorous elements. Such diagnostic information can be generated in pre-tests prior to starting an advertising campaign and thus enable marketers to make informed decisions.

GfK EMO Sensor also provides input at the level of the attention attracted by an advertisement and the personal engagement of the audience. The intensity of emotional experiences is assessed by counting the number of emotions elicited by an advertisement. This intensity score is associated with the electrodermal reaction triggered by a TV ad (see picture validation study D) and can therefore indeed be regarded as a valid indicator of personal engagement: TV ads with higher emotional intensity scores increase arousal and attention in the viewer, thus fostering engagement.
As illustrated by the GfK Advertising Effectiveness Model in the introduction (see Figure 1), engagement has a significant impact on the success of advertising. An analysis across the different campaigns included in the GfK EMO Sensor database shows that the behavioral impact of an ad increases with the intensity of viewers’ emotional experience (level 4 in the advertising model): The higher the level of emotionality, the more likely are respondents to include the advertised brand in their relevant set or even pick it as their choice. Figure 11 shows that this effect can mainly be attributed to the elicitation of positive emotions, whose intensity (just like the intensity of negative emotions) can again be calculated by counting the number of selected positive emotions.

The results from GfK EMO Sensor are also associated with other indicators of advertising effectiveness of the GfK Advertising Effectiveness Model (see Figure 1). A TV campaign for Deutsche Telekom shown all over Europe, featuring the currently well-known casting show candidate Paul Potts, is used to illustrate these relationships. GfK EMO Sensor demonstrates that this TV commercial induces the strongest emotional

Figure 11

Emotional intensity in groups of respondents with different persuasion levels

<table>
<thead>
<tr>
<th>Motivation through TV spots</th>
<th>Total emotional intensity</th>
<th>Emotional intensity with positive valence</th>
<th>Emotional intensity with negative valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>No shift</td>
<td>4.4</td>
<td>3.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Improvement in selection in Relevant Set</td>
<td>5.0</td>
<td>3.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Improvement in selection as First Choice</td>
<td>5.7</td>
<td>4.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>
responses among all ads tested so far. In particular, strong positive emotions such as surprise, pride and passion are elicited (4).

As a consequence of its positive emotionality, the spot leads to a noticeable positive change in brand personality compared to a control group who did not view the spot (level 2: brand associations). The brand is associated more frequently with personality traits such as “enthusiastic”, “open-minded”, and even “vulnerable” – traits that stand in harsh contrast to the negative stereotypes commonly associated with the Telekom brand. Other criteria of success, that belong to level 2 in the advertising model such as brand image, are enhanced as well. Image components such as “treating customers honestly and fairly” or “offering uncomplicated services” are evaluated more favorably than in the control group.

However, emotions are only a necessary, but not sufficient condition for advertising success, as demonstrated by effectiveness criteria in level 3 in the advertising model. There is no improvement in rational brand evaluations of Telekom such as in brand competence (service quality, superiority). Considering that the spot does not provide any direct product information, this comes as no surprise. Emotional aspects of brand attractiveness, in contrast, do show improvement: The advertising campaign succeeds in increasing brand uniqueness and, more importantly, sympathy for Telekom. Additionally, the liking of the spot itself is extremely high compared to other spots in the GfK pre-test database. Finally, brand loyalty (level 4 of the advertising model) is strengthened, as indicated by increased disposition to repeat usage.

Conclusion and future prospects

GfK EMO Sensor is a picture-based, thoroughly validated tool that makes possible differentiated measurement of emotional responses to advertising. The combination of pictures and appropriate emotion labels proved superior in a comprehensive comparison study, by assuring precise, unambiguous understanding of the pictures and facilitating spontaneous, honest answers.

In initial applications, the research tool demonstrated its usefulness in the context of advertising effectiveness research. For instance, the valence of the emotions elicited by a TV commercial could be associated with the intention to buy as well as with other, softer criteria of ad effectiveness (e.g., brand attitude, attitude towards the advertisement). Further applications in market research are currently being explored, for example, in the evaluation of retail stores and new products, as well as in mystery shopping and car clinics (where cars and car accessories together with competitive products are assessed by drivers before or shortly after market launch). GfK EMO Sensor will pro-
vide marketers with better access to the motives underlying consumer behavior. Thus, marketing and market research are supplemented by a reliable measurement tool for evaluating marketing activities and optimizing them from a consumer perspective. Future research will address further relationships. For instance, emotions are expected to increase advertising recall and recognition independently of their valence (level 1 in the GfK Advertising Effectiveness Model, see Figure 1) – a hypothesis that remains to be tested.

Notes

(1) The GfK Advertising Effectiveness Model distinguishes structural and executional catalysts. Examples of structural catalysts are the type of advertising medium (e.g., exhibitions stands tend to create more attention and provide more information than TV ads) and the relevance of the advertised product to the audience (for a family, for instance, a people carrier is of much more interest than a two-seater). The structural catalysts comprise the general framework for communication and cannot be controlled by marketers. They are decisive for the willingness to engage with an advertising medium and are therefore necessary requirements for advertising success. Among the executional catalysts are well-known creative techniques such as use of stimuli of high sensory intensity (colors, loud sounds) or emotional stimuli (surprising twists, humor, physically attractive models; cf. Kroeber-Riel 1993, p.101 ff.). The present article focuses on the latter.

(2) According to Hauth (2007, p.23), "engagement occurs when consumers devote some mental time and effort to the brand or brand communication". He also highlights the close connection between engagement and emotions. For more information on the involvement construct, see Kroeber-Riel et al. (2008) and Hupp (1999).

(3) Moods, attitudes, personality traits, and feelings are closely related to the emotion construct (Scherer 2005, p.703 ff.). In contrast to emotions, moods are more diffuse and longer-lasting affective states, often without apparent cause (e.g., bad mood vs. the emotion of anger). Attitudes, which are highly relevant in brand research, can also have a strong affective component, besides cognitive and behavioral aspects. But again, attitudes are relatively long-lasting and their affective intensity is lower. Similarly, personality traits include emotional aspects and describe relatively stable tendencies of a person to experience certain emotions and moods more frequently (e.g., irritable persons often experience anger). The term feeling, finally, denotes the subjective experience component of emotions and thus is only one element of the emotional response.

(4) The ad was tested using the GfK AD*VANTAGE Multimedia pre-test system in a two-group experimental design, with a control group which was not shown the ad. Thus, by comparing test and control groups, changes in brand image and brand strength can clearly be attributed to the advertisement.
References


