Motivation and Emotion, Vol. 18, No. 1, 1994

Subjectively Salient Dimensions of Emotional Appraisal

Rainer Reisenzein[1,3] and Christine Spielhofer[2]

Four empirical studies of cognitive appraisals in emotion are reported. In studies 1 and 2, a simplified version of the repertory grid method was used to determine subjectively salient dimensions of cognitive appraisal. For a representative sample of 30 emotions, subjects considered pairwise comparisons of remembered eliciting events (study 1) or those typically conducive to the emotions (study 2) and indicated attributes on which the situations differed. The attributes were classified using a category system derived a priori from the theoretical and empirical literature. Some evidence was obtained for the majority of the 25 distinguished potential dimensions of appraisal, and no further dimensions of appraisal were suggested by the data. The most frequently mentioned dimensions — accounting together for 85% of the attributes — were subjective evaluation, causality/agency/responsibility, focus of event, controllability, importance, moral evaluation, stability, social relation positive-negative plus close-distant, self-evaluation, time of event, evaluation of others, intentionality/activity and expectedness. A reduced set of 22 dimensions for which some evidence was obtained in the grid studies was further examined in studies 3 and 4 using a nominal scale analogue of the rating method. It was found that (a) the appraisal dimensions which emerged as the most salient ones in the grid studies tended to be those relevant for the greatest number of emotions, (b) the dimensions were largely statistically independent within the investigated domain of emotions, and (c) they permitted from moderate to good statistical classification of the situations into the emotion categories. Potential limitations of the grid method as well as the issue of the criteria for cognitive appraisals in emotion are discussed.

[1]Free University Berlin, Berlin, Germany.
[2]University of Vienna, Vienna, Austria.
[3]To whom correspondence should be addressed at Department of Psychology, University of Bielefeld, Box 100131, 3350 Bielefeld, Germany.

0146-7239/94/0300-0031$07.00/0

The fundamental tenet of cognitive emotion theorists is that certain types of cognitions, usually called cognitive appraisals (e.g., Lazarus, 1968), play a central role in emotional states. Most cognitive emotion theorists hold that different emotion types are associated with different types of appraisal, and there is also widespread agreement that these emotion-specific appraisals are "composed" out of a limited number of basic components, features, or values on dimensions (for reviews, see Clore, Schwarz, & Conway, 1993; Dalkvist & Rollenhagen, 1989; Lazarus & Smith, 1988; Scherer, 1988). What is still controversial is the *number and identity* of the dimensions of emotional appraisal. Although this issue has occupied emotion theorists since Aristotle, systematic attempts to enumerate and empirically validate the dimensions of appraisal have only been made in recent years. Several theoretical proposals are now available (e.g., Frijda, 1986; Mees, 1985; Ortony, Clore, & Collins, 1988; Roseman, 1984; Scherer, 1984, 1988; Smith & Ellsworth, 1985; Solomon, 1976; Weiner, 1986), as is a set of relevant empirical studies (e.g., Ellsworth & Smith, 1988; Frijda, 1987; Frijda, Kuipers, & ter Schure, 1989; Gehm & Scherer, 1988; Manstead & Tetlock, 1989; Mauro, Sato, & Tucker, 1992; Reisenzein & Hofmann, 1990; Roseman, 1984; 1991; Roseman, Spindel, & Jose, 1990; Smith & Ellsworth, 1985, 1987; Scherer, 1993; Smolenaars & Schutzelaars, 1986/1987; Tesser, 1990; Weiner, 1986).

The present article is a further contribution to this literature. It is a sequel to the study of Reisenzein and Hofmann (1990), in which a new method — a variant of Kelly's (1955) repertory grid procedure — was used for determining dimensions of appraisal that are spontaneously used by people. Before we describe the goals of the present research, the motivation for and the rationale of this method will be summarized.

With few exceptions, the method used in previous empirical studies of emotional appraisals consisted of having subjects rate real, remembered, or imagined eliciting situations for a variety of emotions on scales designed to tap proposed appraisal dimensions. (A dimension was regarded as supported if it contributed to the discrimination between emotions and was largely independent, in a statistical sense, of the further dimensions also considered.) The advantages of this method are that it permits an explicit testing of proposed dimensions, is sensitive to subtle distinctions, and generally leads to results that are well interpretable. It suffers, however, from a potentially crucial disadvantage: It restricts and directs subjects' judgments by means of the choice of scales (Dalkvist & Rollenhagen, 1989; Gigerenzer, 1981; Reisenzein & Hofmann, 1990; Smith & Ellsworth, 1985). Therefore, it is possible (a) that important dimensions of appraisal were overlooked because no scales were included to assess them or (b) that some of the scales measured dimensions of appraisal that are only of minor im-

portance or are even not at all spontaneously used for the appraisal of emotion-eliciting events.

Because of these disadvantages of the rating scale method — which are, of course, not limited to investigations of cognitive appraisals — it is frequently recommended to supplement or even replace this method by multidimensional scaling analyses of similarity judgments or sortings (in the present case, e.g., of hypothetical or remembered emotion-eliciting situations) (see Coxon, 1982; Gigerenzer, 1981). In contrast to ratings on predesigned scales, similarity judgments permit the subjects to use their own dimensions when comparing the objects. However, this method has its own serious disadvantages. Subjects may only attend to the most salient dimensions; they may use different and inconsistent criteria to judge the similarity of different pairs of objects (Smith & Ellsworth, 1985); and the — usually unchecked — representational assumptions of the multidimensional scaling models may be untenable (e.g., Smith & Medin, 1981; Tversky, 1977).

Motivated by these considerations, Reisenzein and Hofmann (1990) proposed an alternative method to determine the dimensions of appraisals that are actually used by people, namely, a simplified version of the repertory grid technique originally developed by Kelly (1955). Subjects were presented with pairwise comparisons of typical eliciting situations of 23 common emotions and were asked to indicate an attribute (i.e., a pair of features) on which the compared situations differed. The resulting attributes were classified — in a primarily inductive, data-driven manner — into several categories, which were then compared with the dimensions of appraisal proposed by various theorists. The grid method shares the advantage of similarity ratings or sortings of not restricting or predetermining the dimensions used by the subjects. In contrast to the latter method, however, the grid procedure was expected both to be sensitive to subtler distinctions and to yield results that are readily interpretable. These expectations were supported. Fourteen categories of attributes could be distinguished, 10 of which qualified as (groups of) potential dimensions of emotional appraisal. The nature and interrelation of the appraisal dimensions suggested by this study was then further clarified in a study using a nominal-scale analogue of the rating scale method (Reisenzein & Hofmann, 1990, study 2). We come back to the results of these investigations in the discussion of the studies reported in the present article.

Based on their findings, Reisenzein and Hofmann (1990) recommended the grid procedure as a valuable adjunct to existing methods in the study of emotional appraisals. In particular, it seems to be a natural complement to the rating scale method used in previous studies, because it yields information on precisely those issues on which this method is silent: first, whether subjects use dimensions of appraisal not considered in existing appraisal models; and second, whether the dimensions proposed in these mod-

els are indeed naturally used by people and, if so, how importantly they feature in subjects' spontaneous construals of emotion-eliciting events. Hence, by combining both methods, it may be possible to arrive at firmer conclusions concerning the dimensions of cognitive appraisal in emotion.

OBJECTIVES OF THE PRESENT RESEARCH

The studies reported in this article are a continuation and extension of the work by Reisenzein and Hofmann (1990). In studies 1 and 2, the grid method was again used. However, several significant improvements were made.

1. In contrast to Reisenzein and Hofmann (1990), the attributes produced by the subjects were categorized using a coding system derived a priori from the theoretical and empirical literature. The development of this coding system was guided by the goals of (a) compiling a comprehensive but nonredundant inventory of previously proposed appraisal dimensions and (b) clarifying ambiguous or unclearly defined dimensions.
2. To reduce the danger that dimensions of appraisal relevant for only a subset of emotions would go undetected, a greater number of emotions (30, compared with the 23 used by Reisenzein and Hofmann) was included, and a greater number of comparisons between emotional situations was used.
3. Two kinds of relevant objects were used in the grid studies: concrete, remembered emotion-eliciting events (study 1) and the perceived typical elicitors of emotions (study 2).
4. To make maximum use of the available information, as well as to avoid problems of classification of complex attributes, *all* dimensions of appraisal that were referred to in a response were coded.

In studies 3 and 4, the potential dimensions of appraisal for which some evidence was obtained in the grid studies were examined further by means of a nominal-scale analogue of the rating method. These analyses are explained in more detail in the Introduction to studies 3 and 4.

STUDIES 1 AND 2

Subjects

The subjects in study 1 were 34 students (25 female; mean age, 26.2 years) of various disciplines at the University of Vienna, who volunteered

to participate in an unspecified "psychological investigation"; those in study 2 were 29 (22 female; mean age, 23.2 years) introductory psychology students at the Free University Berlin who participated in partial fulfillment of their study requirements.

Procedure

As by Reisenzein and Hofmann (1990), a simplified version of Kelly's (1955) grid method was used (see also Fransella & Bannister, 1977; Adams-Webber, 1979). The subjects were asked to compare two emotion-eliciting situations at a time and to indicate an attribute that distinguished between them. Reisenzein and Hofmann (1990) also asked each subject, in accord with the original grid procedure, to rate all emotional situations on all attributes produced by that subject. This step was omitted because it had yielded only limited additional information in the previous study and because it was largely taken care of by the rating procedure used in studies 3 and 4.

Study 1

The participants in study 1 were individually interviewed by the second author. The interview, which lasted from 2 to 4 hr and was conducted in up to two sessions, consisted of two consecutive phases. In phase 1, the subjects were given a list of 30 emotion names and were asked to recall for each of the emotions an episode, preferably one that had recently happened, where they had experienced the respective affect. The recalled emotional situations were characterized by a brief sentence (e.g., "I experienced anger when a friend of mine forgot about our appointment and left me waiting"), which was noted down to serve as a memory aid in the subsequent phase of the study. This second phase differed slightly for subjects 1–20 and 21–34. The first 20 subjects were presented with 20 pairwise comparisons of the emotions. These comparisons were selected, separately for each subject, at random from the 435 possible pairwise comparisons between the 30 emotions, with the restriction that each emotion had to occur at least once. For each of the emotion pairs in turn, the subjects were asked to revisualize the corresponding eliciting situations, to compare them with one another, and to indicate an attribute (i.e., a pair of features) on which the situations differed. It was pointed out that pairs of antonyms, as well as complementary concepts and word pairs denoting simply the presence vs. absence of a quality could be used (cf. Lehrer, 1974). As by Reisenzein and Hofmann (1990), the subjects were encouraged to try to find

different attributes for different emotion comparisons. Once an attribute had been found, the subjects were asked to formulate a sentence that captured the perceived difference(s) between the two situations, such as "Situation A was positive for me; situation B was negative" or "In situation A, I did something wrong; in situation B, someone else did something right." These sentences, which were recorded by the interviewer, constituted the raw data obtained from the investigation.

To check whether infrequent use of some categories by the first 20 subjects was due to an undersampling of particular emotion contrasts, the remaining 14 subjects received a differently sampled set of comparisons. The list of emotions was split into three groups of 10 emotions consisting, respectively, of the positive emotions (including *longing*) and two groups of relatively similar negative emotions (cf. Schmidt-Atzert & Ströhm, 1983; Shaver *et al.*, 1987). This permitted the use of all 45 pairwise comparisons between the situations associated with the emotions within each group, with each subject judging 15 comparisons from each of two emotion groups.

Study 2

In study 2, a questionnaire containing 20 pairwise comparisons between the emotions as well as a detailed instruction was handed out to the subjects. The comparisons were again selected, separately for each subject, according to the method of restricted random sampling used in the first study for subjects 1–20. Rather than to recall a concrete personal experience for each emotion, however, the subjects were now asked to consider the *typical* eliciting situations associated with the two compared emotions and to indicate an attribute that discriminated between these situations. This procedure was based on the assumption that knowledge of appraisal-emotion relationships is a central aspect of people's implicit, naive psychology of emotions (cf. Reisenzein and Hofmann, 1990, study 2) and should therefore also show up if subjects are asked to compare typical eliciting situations. Indeed, the results obtained by this method might even be expected to be clearer because it requires subjects to consider what is common to the situations conducive to a particular emotion; thereby, the naming of features idiosyncratic to specific situations (study 1; see also Reisenzein and Hofmann, 1990, study 1) may be avoided. However, because of the more abstract nature of the task, it was deemed necessary to illustrate the procedure with a realistic example of an appraisal dimension. *Time of event*, a category that emerged with moderate frequency in both Reisenzein and Hofmann (1990) and the present study 1, was chosen for

this purpose. The results of study 2 for this category must therefore be regarded with caution.

Selection of Emotions

The emotions were selected according to both theoretical and empirical criteria, i.e., we included emotions that have been accorded central importance by contemporary emotion theorists or are regarded as typical examples of the category "emotion" by laypeople (see, e.g., Fehr & Russell, 1984; Schmidt-Atzert, 1981). The resulting list of 30 emotions, which is a compromise between these two selection criteria, can be regarded as a fairly representative sample of the emotional spectrum (cf., e.g., Schmidt-Atzert & Ströhm, 1983; Shaver *et al.*, 1987). For example, it includes at least one member of 19 of the 25 groups distinguished by Shaver *et al.* (1987) in a hierarchical cluster analysis of similarity sortings of 135 typical emotion terms. The following emotions were included (the original German names are listed in parentheses): admiration (Bewunderung), anger (Ärger), anxiety/fear (Angst/Furcht), apprehension (Besorgnis), contempt (Verachtung), contentment (Zufriedenheit), despair (Verzweiflung), disappointment (Enttäuschung), discontentment/dissatisfaction (Unzufriedenheit), disgust (Ekel), embarrassment (Verlegenheit), envy (Neid), gloating (Schadenfreude), gratitude (Dankbarkeit), hate (Haß), hope (Hoffnung), indignation (Empörung), joy (Freude), guilt (Schuld), hopelessness (Hoffnungslosigkeit), jealousy (Eifersucht), loneliness (Einsamkeit), longing (Sehnsucht), love (Liebe), pity (Mitleid), pride (Stolz), regret/remorse (Reue), relief (Erleichterung), sadness (Trauer), and shame (Scham). Twenty-two of these emotions had already been used by Reisenzein and Hofmann (1990); the remaining eight (apprehension, despair, indignation, hate, longing, admiration, gloating, and contentment) were newly added.

Coding System

The coding system was specifically developed for the present studies. It consists of 39 categories. Twenty-five of these represent potential dimensions of cognitive appraisal, 13 refer to aspects of emotions other than appraisals (including emotion names), and the last category ("unclassifiable") was reserved for idiosyncratic attributes.

Appraisal Categories

The categories for (potential) appraisal dimensions were compiled from the relevant theoretical and empirical literature. Each dimension proposed in at least one of the sources that were searched was initially regarded as a possible candidate, provided that it represented a potential judgment of (an aspect of) an emotion-eliciting situation or object. This criterion resulted, for example, in the exclusion of *attentional activity* (Smith & Ellsworth, 1985) as this dimension is more adequately construed as a kind of action tendency (Frijda *et al.*, 1989; Lazarus & Smith, 1988). For each of the resulting dimensions, a brief characterization was compiled consisting of explanatory quotations from the respective author, possible synonyms and references to dimensions proposed by others (if listed), and, if available, questions or items used for the measurement of the dimensions. An attempt was then made to clarify dimensions that were ambiguously or unclearly described, and in some cases dimensions were slightly redefined. For example, the dimensions of *stability* and *controllability*, originally proposed by Weiner (e.g., 1982) as dimensions of causes of events, were redefined to refer directly to the emotion-eliciting events, thus becoming largely equivalent to, respectively, Frijda's (1986) dimension of modifiability-finality and Roseman's (1984) power (cf. also Scherer, 1984, 1988). Subsequently, conceptually identical or highly similar dimensions were grouped together [*causality/agency* and *responsibility* were also grouped together because, although conceptually distinct (e.g., McGraw, 1987), these dimensions are usually strongly associated empirically and were difficult to keep apart in the coding]. An exception was made, however, for *evaluative* dimensions (see below), which were not combined into larger categories even though this would have been possible in some cases. The reason was that we believe, in line with both older (e.g., Stumpf, 1899) and several contemporary cognitive emotion theorists (e.g., Arnold, 1970; Lyons, 1980; Mees 1985; Ortony *et al.*, 1988; Solomon, 1976) that evaluations are central to emotional states (see also Reisenzein & Schönpflug, 1992). Therefore, we wanted to give particular attention to these dimensions. As a consequence, the coding system has a somewhat higher resolution for evaluative than for nonevaluative dimensions.

The resulting 25 potential dimensions of cognitive appraisal are described in detail in the Appendix. The following comments are therefore restricted to a description of the general structure of the appraisal coding system and some of its peculiarities.

On a most general level, the 25 potential dimensions of appraisal can be divided into those that are intrinsically evaluative in character — in particular, subjective, interpersonal, and moral evaluation, self- and other

evaluation, evaluation of relationship positive–negative and close–distant, superior–inferior, and importance — and those that are not (e.g., controllability or time of event).

The evaluative dimensions can be further subdivided, according to the object of appraisal, into at least two subgroups: evaluations of states of affairs (including stable conditions, events and actions) and evaluations of objects. *Object evaluations* (e.g., Frijda, 1986; Ortony *et al.*, 1988) were differentiated into self- and other evaluations. *Evaluations of states of affairs* were divided, following Solomon (1976), into three subtypes differing with regard to the standard of evaluation on which they are based. *Subjective evaluations* capture evaluations of an eliciting state of affairs based on personal goals and preferences that may be quite idiosyncratic and do not concern social or moral norms (goal conduciveness can be regarded as a special subform of this category); *interpersonal evaluations* reflect the emotion experiencer's concern with how the state of affairs in question might be evaluated by significant others; and *moral evaluations* comprise evaluations of states of affairs with regard to internalized moral or ethical norms that are (typically, at least) seen as "objectively" valid (e.g., Neppl & Boll, 1991). The distinction between moral and nonmoral evaluations is a common one in the recent literature on appraisals, although the former are sometimes restricted to judgments of fairness or legitimacy. The category of interpersonal evaluations is less frequently found (e.g., Solomon, 1976) but comparable dimensions of appraisal were suggested, for example, by Krech and Crutchfield (1958), Scherer (1984, 1988), and Manstead and Tetlock (1989). Interpersonal evaluations also figure prominently in the literature on shame (e.g., Taylor, 1987).

Three further evaluative dimensions pertain to aspects of *social relationships*. Two of these (*relationship positive–negative* and *close–distant*) were suggested by the results of Reisenzein and Hofmann [1990; cf. also Solomon's (1976) "intersubjectivity" judgments]; the third (*superior–inferior*) was suggested by Krech and Crutchfield (1958) as a basic dimension of "interpersonal" emotions [cf. also Solomon's (1976) "personal status"]. Although these dimensions could perhaps be subsumed under the other and self-evaluation categories, respectively, or be construed as a combination of these categories, we decided to keep them separate because of the limited attention that appraisals pertaining specifically to social relations have so far received in empirical studies.

Turning to the nonevaluative categories, *intentionality/activity* emerged as a category in Reisenzein and Hofmann (1990; there termed activity); although related to *causality/agency*, this dimension is conceptually distinct (e.g., Weiner, 1986; Frijda, 1986) and was therefore treated separately. The *certainty/probability* dimension suggested by a number of authors (e.g., Fri-

jda *et al.*, 1989; Roseman, 1984; Smith & Ellsworth, 1985) was split into certainty concerning the eliciting event and predictability of its consequences; *stability* (Weiner, 1982) was redefined as mentioned above, i.e., as modifiability–finality in Frijda's (1986) sense. *Controllability* was defined as comprising both the controllability of the eliciting event (if it had not yet occurred) and/or of its consequences. Finally, the dimension *focus*, which reflects the emotion experiencer's belief as to who is primarily affected by an eliciting state of affairs (oneself or someone/something else), deserves to be mentioned separately because it has been so far neglected in most empirical studies; however, it seems to be of central importance for the distinction between self-related and "empathic" emotions, i.e., those which occur in reaction to the perceived fate of other people, such as pity or gloating (cf. Markus & Kitayama, 1991).

For additional information on these and the further appraisal categories, the reader is referred to the Appendix and the references cited there.[4] Note also that the structural models of appraisals proposed by various recent authors (cf. the Introduction) can, for the most part, be reconstructed as *subsets* of the present set of appraisal dimensions. By selecting appropriate subsets of dimensions, it would therefore be possible to compare the various proposed models on a variety of criteria (cf. the General Discussion). However, in the present studies we refrained from such model comparisons, because we think that the important, actually used appraisal dimensions should be determined independently of who proposed them.

Nonappraisal Categories

The remaining 13 categories were suggested in part by the literature and the previous study by Reisenzein and Hofmann (1990); partly they emerged during the coding of the data from study 1. The main reason for their inclusion was to reduce the number of attributes that would otherwise have had to be coded as unclassifiable, thereby facilitating the decision of whether any further potential dimensions of appraisal, not considered in the coding system, were suggested by the data. The categories comprise the following. (a) Noncognitive mental states that are frequently associated with, and are regarded by some as central components of emotions, namely *action tendencies* or *performative desires* (i.e., desires to do something, cf. Frijda, 1986) *wishes* or *nonperformative desires* (i.e., desires that something

[4]We are aware that the category system may contain a number of biases that reflect our theoretical predilections. Readers who miss a dimension they regard as important are asked to consult the examples of attributes listed in the Appendix to decide whether they might be taken as representing a further, or somewhat differently defined, appraisal dimension.

should be or become in a certain way without one's own intervention, cf. Peters, 1970), and *perceptions of physiological symptoms or general arousal* (cf. Schachter, 1964). Two specific types of action tendencies (*attentional activity* and *control of emotion*) were separately coded, the first because it was assigned special importance for emotions and was empirically supported by Smith and Ellsworth (1985, 1987); the second, which has been distinguished as a particular form of coping strategy by Lazarus (e.g., 1991), because it was mentioned with some frequency by our subjects. (b) Various basic features of emotional experience, namely, *intensity, duration, and the direction of emotion* (e.g., angry at self versus others). (c) A number of perceived situational differences that were mentioned with some frequency but are too specific to be regarded as appraisal dimensions (*specific situations, emotion caused by person versus inanimate object, involved person known versus unknown,* and *situation rationally explainable*). (d) *Emotion names* as well as an *unclassifiable* category.

Coding Procedure

Using the coding system just described, about one-third of the subjects' responses were coded jointly by the two authors and about one-third separately by each one. To make maximum use of the information contained in the data as well as to avoid problems of classification of complex attributes (i.e., those involving simultaneous contrasts between more than two features) all of the categories identifiable in a subject's response were coded; also, a category was regarded as present even if only one member of a feature pair indicating that category (e.g., "unjust" for the moral evaluation category) was mentioned. This procedure resulted in up to four codings per response, although four codes occurred only twice. To illustrate, the response "situation A is positive for me, results from the effort of others and concerns me; situation B is negative and concerns others" was coded as *subjective evaluation of the eliciting state of affairs* (positive–negative), *causality/agency* (results from the effort of others), and *focus* (concerns me–concerns others).

Results

In study 1, data from 790 comparisons between emotional situations (on average 23.2 per subject) were obtained; in study 2, from 578 comparisons (M = 19.9 per subject). Missing responses (30 in study 1 and 2 in study 2) were due to the fact that subjects either could not recall an emotional episode for one or more of the emotions (study 1) or could not think of a discriminating attribute for a comparison (both studies). Across both studies, each

of the possible 435 emotion comparisons occurred on average 3.14 times (SD = 2.3), with 395 (91%) of the comparisons occurring at least once. Each emotion appeared on average 91.2 times as a member of a comparison, ranging from 75 to 102 occurrences. The total number of coded attributes was 1029 (on average 30.3 per subject) in study 1 and 882 (30.4 per subject) in study 2. Hence, 239 (23%) of the attributes from study 1 and 304 (34%) of those from study 2 were second to fourth codings, documenting that a substantial number of the responses involved more than one attribute.

Coding Reliability

To estimate the reliability of the codings, the whole data set from study 1 was recoded by the second author 7 months after the first coding. The chance-corrected proportions of agreement between the two codings for the individual categories, as expressed by conditional kappa (Cohen, 1960), are reported in Table I, column 3. These scores are the average of the conditional kappa values obtained if once the first and then the second coding was used as the comparison standard (cf. Hubert, 1977). (Kappas were computed for all categories with greater than zero frequencies, but it is clear that those based on very low frequencies are rather imprecise estimates of reliability.) Considering the large number of categories, the reliabilities for the majority of them can be regarded as acceptable, particularly for the appraisal categories, which were of main interest. The mean conditional κ for the 36 categories with frequencies > 0 was .67, and .70 for the 23 appraisal categories with greater than zero frequencies. Sixteen of the latter categories had κ's $\geq .70$, and all but two (difficulty and anticipated effort) $\geq .50$.

Effect of Set of Comparisons

To decide whether the different sets of comparisons for subjects 1–20 and 21–34 in study 1 had an effect on the results, the frequency distributions of the categories for the two groups were compared (.01 was added to cells with zero frequencies). There was no significant difference [$\chi^2(N = 1029, df = 38) = 42.2.$, ns]. Therefore, the data from the two subgroups of study 1 were collapsed.

Differences Between Study 1 and Study 2

Columns 4–5 and 6–7 in Table I contain the category frequencies and percentages of the 39 categories for studies 1 and 2, respectively. Sta-

Table I. Frequency Distribution of Categories for Studies 1 and 2

Category name	Number	Reliability (κ)[a]	Study 1		Study 2	
			F	%	*F*	%
Appraisal						
Subjective evaluation	1	.77	115	11.2	192	21.8*
Goal conduciveness	2	—[b]	0	0.0	1	.1
Interpersonal evaluation	3	.56	9	.9	19	2.2
Moral evaluation	4	.80	21	2.0	40	4.5
Self-evaluation	5	.75	26	2.5	25	2.8
Evaluation of others	6	.70	22	2.1	26	2.9
Social relation: positive–negative	7	.57	16	1.6	10	1.1
Social relation: superior–inferior	8	.79	15	1.5	7	.8
Social relation: closeness–distance	9	.73	12	1.2	19	2.2
Importance	10	.84	87	8.5	21	2.4*
Time	11	.93	28	2.7	82	9.3*
Suddenness	12	.75	8	.8	1	.1
Expectedness	13	.75	11	1.1	29	3.3
Familiarity	14	.96	12	1.2	5	.6
Certainty/probability	15	.75	13	1.3	15	1.7
Predictability of consequences	16	.62	4	.4	4	.5
Stability	17	.59	39	3.8	19	2.2
Controllability	18	.83	78	7.6	31	3.5*
Causality/agency/responsibility	19	.73	94	9.1	70	7.9
Intentionality/activity	20	.58	34	3.3	13	1.5
Focus	21	.77	62	6.0	59	6.7
Anticipated effort	22	.47	14	1.4	4	.5
Focality–globality	23	.83	25	2.4	6	.7
Difficulty	24	.00	3	.3	5	.6
Interestingness	25	—	0	0.0	0	.0
Nonappraisal						
Specific situations	26	.36	21	2.0	3	.3
Emotion caused by person vs. object	27	.55	13	1.3	16	1.8
Involved person known–not known	28	.61	13	1.3	0	.0
Situation rationally explainable	29	.93	6	.6	4	.5
Nonperformative desires (wishes)	30	.58	17	1.7	22	2.5
Performative desires (action tendencies)	31	.80	57	5.5	32	3.6
Attentional activity	32	.50	2	.2	1	.1
Control of emotion	33	.75	12	1.2	0	.0
Direction of emotion	34	.51	21	2.0	1	.1*
Duration of emotion	35	.66	36	3.5	6	.7*
Intensity of Emotion	36	.57	33	3.2	1	.1*
Physiological symptoms/arousal	37	.86	15	1.5	13	1.5
Emotion terms	38	—	0	0.0	56	6.3*
Unclassifiable	39	.45	35	3.4	24	2.7
Total		.67	1029	100.0	882	100.0

[a]Determined for the data from study 1.
[b]Index not computable.
*Category frequencies of studies 1 and 2 differ significantly at $\alpha = .05/39$ (χ^2 test, $N = 1911$, df = 1).

tistical comparisons between the frequency distributions obtained in the two studies revealed significant differences both if all categories were taken into account [$\chi^2(N = 1911, df = 38) = 336.5, p < .001$] and if only the appraisal categories were considered [$\chi^2(N = 1451, df = 24) = 175.9, p < .001$]. Posteriori chi-square tests for the individual categories, with α adjusted to .05/39, revealed that two appraisal categories (subjective evaluation and time of event) and one nonappraisal category (emotion terms) had significantly higher relative frequencies in study 2 than in study 1, whereas the reverse was true for two other appraisal categories (importance and controllability) and three nonappraisal categories (direction of emotion, duration of emotion, and intensity of emotion; cf. Table I).[5] The higher relative frequencies of emotion terms and time of event found in study 2 are most likely artifacts caused by procedural differences. As to *emotion terms*, although the subjects in both studies were informed that emotion names did not count as valid responses, in study 1 the interviewer could reject such responses if they were nevertheless given, which was not possible in study 2, where the responses were obtained in written format. The higher relative frequency of *time of event* observed in study 2 was most likely due to the fact that, as mentioned, this appraisal category was used to illustrate the procedure. In contrast, the remaining differences between the studies can be plausibly attributed to the personal and concrete vs. impersonal and abstract nature of the judged objects. In particular, it is plausible that basic phenomenal features of the subjective experience of emotion, such as the *intensity*, *duration*, and *direction* of emotion are more salient when personal experiences of emotional situations (study 1) than when abstract situation types (study 2) are compared; and it is also plausible that people in the former case are more concerned about the *importance* of the eliciting events and about their possibilities for *personal control*. Finally, *subjective evaluation* of events, the dominant category in both studies, became apparently even more salient when the typical situations conducive to emotions, rather than concrete personal experiences were compared.

Despite the significant differences between the two studies, the magnitudes of the obtained differences were in general moderate (cf. Table I), particularly for the appraisal categories, for which a rank-order correlation of .75 between the two studies was obtained. Because the obtained differences did not require that substantially different conclusions be drawn depending on which data set was considered, we decided to collapse the data from both studies and to restrict the further discussion to the results for

[5]Note that the chi-square tests could be biased to some degree because the frequency data are based on multiple responses from the same subjects (see e.g., Garner & Hartmann, 1984). However, we feel reasonably certain that, due to the α adjustment, false-positive conclusions could be avoided.

Table II. Frequency of Categories for Pooled Data

Category name	Number	Reliability (κ)	*F*	% appraisals	%
Appraisal					
Subjective evaluation	1	.77	307	21.2	16.1
Causality/agency/responsibility	19	.73	164	11.3	8.6
Focus	21	.77	121	8.3	6.3
Time[a]	11	.93	110	7.6	5.8
Controllability	18	.83	109	7.5	5.7
Importance	10	.84	108	7.4	5.7
Moral evaluation	4	.80	61	4.2	3.2
Stability	17	.59	58	4.0	3.0
Self-evaluation	5	.75	51	3.5	2.7
Evaluation of others	6	.70	48	3.3	2.5
Intentionality/activity	20	.58	47	3.2	2.5
Expectedness	13	.75	40	2.8	2.1
Social relation: closeness–distance	9	.73	31	2.1	1.6
Focality–globality	23	.83	31	2.1	1.6
Interpersonal evaluation	3	.56	28	1.9	1.5
Certainty/probability	15	.75	28	1.9	1.5
Social relation: positive–negative	7	.57	26	1.8	1.4
Social relation: superior–inferior	8	.79	22	1.5	1.2
Anticipated effort	22	.47	18	1.2	.9
Familiarity	14	.96	17	1.2	.9
Suddenness	12	.75	9	.6	.5
Predictability of consequences	16	.62	8	.6	.4
Difficulty	24	.00	8	.6	.4
Goal conduciveness	2	—	1	.1	.1
Interestingness	25	—	0	.0	.0
Appraisal categories		.70	1451	100.0	75.9
Nonappraisal				% nonappraisals	
Performative desires (action tendencies)	31	.80	89	19.3	4.7
Unclassifiable	39	.45	59	12.8	3.1
Emotion terms	38	—	56	12.2	2.9
Duration of emotion	35	.66	42	9.1	2.2
Nonperformative desires (wishes)	30	.58	39	8.4	2.0
Intensity of emotion	36	.57	34	7.4	1.8
Emotion caused by person vs. object	27	.55	29	6.3	1.5
Physiological symptoms/arousal	37	.86	28	6.1	1.5
Specific situations	26	.36	24	5.2	1.3
Direction of emotion	34	.51	22	4.8	1.2
Involved person known–not known	28	.61	13	2.8	.7
Control of emotion	33	.75	12	2.6	.6
Situation rationally explainable	29	.93	10	2.2	.5
Attentional activity	32	.50	3	0.7	.2
Nonappraisal categories		.63	460	100.0	24.1
All categories		.67	1911		100.0

[a]Inflated due to study 2 data (see text); for study 1, time is 3.5% of all appraisal codings.

the pooled data. The pooled frequencies of the categories are presented, in decreasing order of frequency, in Table II.

Results for Appraisal Categories

A general problem of interpreting the data obtained in the present investigation concerns the question of when a category should be regarded as supported by the results. As by Reisenzein and Hofmann (1990), a rather liberal criterion was adopted: Any category that was mentioned with some frequency and was coded with acceptable reliability was considered as having received *some degree of* support. Using this criterion, the results listed in Table II can be interpreted as lending some degree of support to all potential dimensions of appraisal except *interestingness*, which was never coded; *goal conduciveness*, which was coded only once; and *difficulty*, which was observed only eight times and was completely unreliable. The support obtained for some of the remaining dimensions was, however, fairly weak, in particular for *predictability of consequences* and *anticipated effort*, both of which were coded infrequently and with low reliabilities. Doubts based on relatively low reliabilities ($\kappa < .60$) could also be raised with regard to *stability*, *intentionality/activity*, *interpersonal evaluation*, and *social relation positive–negative*. It should be noted, however, that — in contrast to difficulty, predictability, and anticipated effort — the disagreements on these categories were not random, but mostly restricted to related categories (stability was confused with duration of emotion, intentionality/activity with causality/responsibility, interpersonal evaluation with moral evaluation, and social relation positive–negative with social relation close–distant).

As mentioned, we take evaluations to be of central importance to emotional states. In line with this assumption, *subjective evaluation* was the dominant appraisal category in both studies (21.2% of all appraisal codings), and all further evaluative categories except goal conduciveness were also observed, although with much lower frequencies: *importance* (7.5%), *moral evaluation* (4.2%), *self-evaluation* (3.5%), *evaluation of others* (3.3%), *social relation close–distant* (2.1%), *interpersonal evaluation* (1.9%), *social relation positive–negative* (1.8%), and *superior–inferior to others* (1.5%). (It should be noted that the subjective evaluation category is probably somewhat inflated relative to the other evaluative categories because it was also coded whenever the standards of an evaluation could not be determined.) Together, these evaluative categories accounted for nearly half (47%) of all appraisal codings. In addition, inasmuch as *desires* can be regarded as evaluative mental states in a wider sense of this term (cf. Reisenzein & Schönpflug, 1992), it is noteworthy that nonperformative desires (wishes:

8.4% of all nonappraisal codings) and performative desires (action tendencies: 19.3%) were the most frequent nonappraisal categories (not counting unclassifiable, emotion terms, and duration of emotion; cf. Table II). Actually, the total frequency of action tendencies is even somewhat higher (22.6%) because *control of emotion* (2.6%) and *attentional activity* (0.7%) can be regarded as special kinds of action tendencies.

However, it is clear from Table II that several nonevaluative (i.e., not intrinsically evaluative) differences between emotional situations were also salient to subjects, most importantly, *causality/agency/responsibility* (11.3%), *focus of event* (8.3%), and *controllability* (7.5%). In fact, these categories were mentioned more frequently than any single explicitly evaluative category except subjective evaluation. [The relatively high frequency of *time of event* (7.6%) was due mainly to the study 2 results, which, as mentioned, are probably an artifact. This dimension accounted for 3.5% of the appraisal codings in study 1.] *Stability* (4.0%) and *intentionality/activity* (3.2%) were next in frequency, but both categories had comparatively low reliabilities. *Expectedness* (2.8%), *focality–globality* (2.1%), *certainty/probability* (1.9%), *familiarity* (1.2%), and *suddenness* (0.6%) were still less frequent, but fairly reliable.

Finally, we reexamined the 59 (3.1%) attributes in the unclassifiable category to see whether they suggested any further potential dimensions of appraisal. The results were negative.

Discussion

We first compare the present findings with those obtained in the previous grid study by Reisenzein and Hofmann (1990, study 1) and then discuss possible reasons for the lack of evidence for some of the appraisal dimensions proposed in the literature.

In Reisenzein and Hofmann's (1990) grid study, evidence was obtained for some 10, partly heterogeneously defined, categories of appraisal dimensions. The coding system used in the previous study was coarser (i.e., several of the dimensions distinguished in the present studies were grouped together in a single category), and some categories were somewhat differently defined (e.g., responsibility was separated from causality/agency and grouped together with moral evaluation). However, it is possible to obtain an acceptable approximation to the previous coding system by combining several of the present categories. Focus, importance, controllability, and intentionality/activity were defined nearly identically in both studies; Reisenzein and Hofmann's valence category is covered by the present categories of subjective, interpersonal, and self-evaluation; their temporal aspects

category had time of event and stability (*unique–lasting situation*) as subcategories; the present expectedness and familiarity categories were joined in a single category in the former study; Reisenzein and Hofmann's causality/agency and responsibility/moral evaluation categories are covered by the present categories of causality/agency/responsibility and moral evaluation; and their social relationship aspects category is covered by social relation positive–negative, close–distant, inferior–superior, and (partly) evaluation of others. When the present categories were combined as indicated to conform as closely as possible to those used in the previous study, and the unbiased frequency estimate for time (from study 1) was used, the rank order of the categories turned out to be identical to that obtained in the previous study except for focus and controllability, both of which were less frequently observed in Reisenzein and Hofmann's (1990) study. The lower frequency of focus was apparently due to the fact that each attribute received only one coding in the previous study and focus was often mentioned as a secondary attribute; when the attributes listed in Table 2 of Reisenzein and Hofmann (1990) were recoded for focus, allowing for multiple codings, this category became nearly as salient as in the present investigations.

No support was obtained by Reisenzein and Hofmann (1990) for a number of other appraisal dimensions proposed in the literature, in particular goal conduciveness (goal–path obstacle) (Scherer, 1984, 1988; Smith & Ellsworth, 1985), anticipated effort (Smith & Ellsworth, 1985), focality–globality, interestingness, self-esteem (Frijda, 1987; Frijda *et al.*, 1989), and difficulty (Smith and Ellsworth, 1987). These negative results were replicated here for goal conduciveness, interestingness, difficulty and anticipated effort (which received only very weak support). In contrast, some evidence for focality–globality and self-esteem was now obtained.

The lack of evidence in the previous investigation for *focality–globality*, defined as the degree to which the emotion-eliciting event is perceived as something concrete and specific vs. something unspecific or diffuse (cf. the Appendix), can be attributed post hoc to the fact that the subjects were asked to compare highly typical scenarios for the various emotions which, as supported by the results of study 4, always described specific or focal eliciting events. The lack of support for Frijda's *self-esteem* was apparently due to the fact that this dimension was interpreted too literally in the original study [cf. Frijda *et al.* (1989): "Did the situation decrease or enhance your self-esteem?"]; if reinterpreted as self-evaluation, as in the present study, we found that it could be detected with some frequency (13) among the attributes obtained in the original investigation.

Interestingness, defined as the judgment of how interesting the event or situation is, may be relevant only, as suggested by Frijda (1986, 1987),

for interest and related emotions (if these mental states are emotions at all) that were not included in the present study. *Perceived difficulty* is, as noted by Smith and Ellsworth (1987, p. 477) themselves, probably relevant only in exam or similar achievement situations. If so, however, it seems too situation-specific to us be counted as a dimension of emotional appraisal (see also Weiner, 1986).

Goal conduciveness, defined as the degree to which the emotion-eliciting event is perceived as furthering or hindering one's goals or plans (cf. the Appendix), was originally proposed by Scherer (1984) and subsequently adopted in modified form (goal–path obstacle) by Smith and Ellsworth (1985; cf. also Frijda *et al.*, 1989). As in the previous rating scale studies by Smith and Ellsworth (1985) and Frijda *et al.* (1989), we obtained no evidence for the independent status of a goal-conduciveness dimension (both Frijda *et al.* and Smith and Ellsworth found that this dimension was highly correlated with pleasantness). Although numerous attributes coded in the subjective evaluation category referred to a personal goal or wish having been achieved versus not achieved [e.g., success–failure, I get something–I lose something, goal reached–not reached, gain–loss (see also Reisenzein and Hofmann, 1990, Table I)], explicit references to something as standing in the way to one's goals (Smith and Ellsworth, 1985) or as hindering or furthering the attainment of one's goals or aims (Gehm & Scherer, 1988; see also Scherer, 1993) were not made by our subjects. It is possible, however, that a number of the attributes grouped into the subjective evaluation category were based on goal-conduciveness appraisals, even though the standards of evaluation were not explicitly mentioned.

Finally, *anticipated effort* was proposed by Smith and Ellsworth (1985) as a replacement for the activation or arousal dimension that frequently emerged in previous scaling studies of emotion terms and facial expressions of emotions. It was intended to represent Cannon's (1929) concept of fight vs. flight. However, neither arousal nor fight vs. flight is a prima facie dimension of cognitive appraisal, and in fact, Smith and Ellsworth (1985) reconceptualized this dimension as referring to the person's *considerations* concerning the degree of effort needed to deal with the situation, thus making it similar to the (outcome of the) process of effort calculation hypothesized in attributional models of achievement motivation (e.g., Kukla, 1972; Meyer, 1973). Anticipated effort emerged as a separate factor in the studies by Smith and Ellsworth (1985, 1987) and Frijda *et al.* (1989, study 1) and was found to contribute significantly to emotion differentiation. Furthermore, Smith (1989) reported a significant correlation between ratings on this dimension and heart rate during the imagery of various emotional events. Although these findings seem to support Smith and Ellsworth's (1985) hypothesis that anticipated effort plays an important role in emo-

tional states, the lack of evidence for this dimension found by Reisenzein and Hofmann (1990), and the very weak support obtained in the present studies, put this assumption into doubt. In our view, it is unlikely that the lack of evidence for this dimension was due to the failure to include emotions for which it may be particularly relevant (e.g., frustration, challenge, or determination), because several of the included emotions were found to have from moderate to high values on this dimension in studies 3 and 4 (e.g., hopelessness, despair, loneliness, sadness, guilt, hate, and jealousy in study 3).

Summary

(1) Some support was obtained for the majority of potential appraisal categories, and the results of Reisenzein and Hofmann (1990) were largely replicated. (2) In terms of relative salience — as indicated by the rank order of frequencies — the most important dimensions were, in this order, subjective evaluation, causality/agency, focus, controllability, importance, moral evaluation, stability, social relationship positive–negative plus close–distant (3.9%; cf. studies 3 and 4), self-evaluation, time of event (study 1 estimate), evaluation of others, intentionality/activity, and expectedness; together, these 14 dimensions accounted for about 85% of all appraisal codings. None of the appraisal theories that have been proposed (cf. the Introduction) includes precisely these high salience dimensions, but there is reasonable agreement with several of these theories, in particular those of Frijda (1986; Frijda *et al.*, 1989), Ortony *et al.* (1988; see also Mees, 1985), and Weiner (1982; 1986; see also Brown & Weiner, 1984). (3) No further appraisal dimensions were suggested by the data. (4) No support was obtained in both the present studies and the previous investigation by Reisenzein and Hofmann (1990) for interestingness, difficulty, and goal conduciveness, and evidence for anticipated effort was weak. The lack of evidence for interestingness may however have been due to the fact that emotions for which this dimension is thought to be particularly relevant were not included, and goal conduciveness appraisals may have been underlying some of the attributes coded in the subjective evaluation category.

STUDIES 3 AND 4

The 22 potential dimensions of appraisal for which some evidence was obtained in studies 1 and 2 were examined further in studies 3 and

4, in which a nominal-scale analogue of the rating scale method (e.g., Smith and Ellsworth, 1985; Frijda *et al.*, 1989) was used (cf. Gehm & Scherer, 1988; Reisenzein and Hofmann, 1990). These additional studies were conducted because several of the included appraisal dimensions have not been empirically investigated before in a rating study, and no prior study included all of these dimensions or the same emotions. The aim of the studies was to further clarify and validate the (reduced) set of appraisal dimensions by determining (a) the degree of statistical redundancy vs. independence of the dimensions, (b) the capacity of the dimensions to discriminate between emotions, and (c) the typicality of dimension values for the different emotions as well as the generality of the dimensions across emotion (see below). In study 3, subjects were asked to judge remembered emotion-eliciting situations on scales intended to assess the dimensions suggested by the previous studies, whereas in study 4, 460 emotion-eliciting scenarios collected in a previous study (Reisenzein and Hofmann, 1993) were "expert coded" on the same scales.

Method

Subjects and Procedure

Study 3. For economic reasons, an attempt was made to secure once more the participation of the subjects of study 1. However, only 22 of these were available or willing to participate; therefore 11 additional subjects from the same subject pool (students at the University of Vienna) were added. Three subjects, two of them new, were excluded from the analysis because preliminary data screening revealed that their response profiles were extremely deviant and there were grounds for suspecting that they answered the questionnaires in a very superficial manner. Twenty of the remaining 30 subjects were female; the mean age was 27.3 years. Subjects who had already participated in study 1 were asked to recall and rate the situations originally reported for the various emotions, whereas subjects who newly participated were asked to first recall and then rate corresponding experiences. The ratings were made using a questionnaire containing items designed to assess the 22 potential appraisal dimensions for which some evidence was obtained in the first two studies (all except *goal conduciveness, interestingness*, and *difficulty*). Due to the large number of dimensions, only one item was formulated for each one; these items are reproduced in the Appendix. Each item was to be answered on a nominal scale comprising three categories. The categories

represented, respectively, (a) the presence of an appraisal component (i.e., one value of the dimension, such as *important, positive*); (b) the absence of that appraisal component or — in the case of bipolar dimensions — the presence of its opposite (e.g., *unimportant, negative*); and (c) cases where none of these response options seemed to apply, that were undecidable, or where the dimension was inapplicable or irrelevant (such as when the quality of relationship had to be judged in a situation which involved no interaction with others). Nominal scales were used because, as they require less refined distinctions, they are less time-consuming to answer than rating scales; because they avoid the problem of possible nonlinear relations between emotions and appraisal dimensions (cf. Neter & Wasserman, 1974); and because we believed that they permit a clearer distinction of appraisal components relevant (categories a or b) vs. not relevant (category c) for each emotion.

Study 4. In study 4, 460 emotion-eliciting scenarios obtained by Reisenzein and Hofmann (1993) through an interview technique were "expert coded" by the authors and Thomas Hofmann (about one-third by each) on the appraisal scales. The 460 scenarios consist of 20 descriptions of eliciting situations for each of the 23 emotions examined by Reisenzein and Hofmann (1990); as mentioned, 22 of these (all except surprise) were also included in the present studies 1–3. As described in more detail by Reisenzein and Hofmann (1993), the scenario descriptions included only appraisal-relevant situational information, whereas all other features from which emotions could be potentially inferred were eliminated. We expected that the coders, being familiar with the theoretical definitions of the dimensions, would use the scales in a more valid and reliable manner than the subjects in study 3. After initial practice of the coding system using different situations, the 460 scenario were coded in random order. The coders were, of course, blind to the emotions portrayed in the scenarios. The task of the coders was to adopt the perspective of the story protagonist and to decide how the situation was appraised by him or her on the 22 dimensions. For purposes of reliability estimation, 50 randomly selected situations coded by the second author (who was completely unfamiliar with the scenarios) were recoded by the first author. Average proportion agreement for the 22 categories was .76 (.64 if corrected for chance). Comparatively low agreements occurred only for the dimensions suddenness (proportion agreement, .54; chance-corrected, .31), familiarity (.58, .37), predictability (.60, .40), stability (.56, .34), controllability (.60, .40), and anticipated effort (.62, .43). For the remaining dimensions, the average proportion agreement was .82 (.74 corrected for chance).

Results and Discussion

Relations Among Appraisal Scales

As mentioned in the description of the coding system, the potential dimensions of appraisal were intended to be *conceptually* distinct; specifically, no dimension was to be in an obvious way definable in terms of others. However, even if this goal was reached, it is still possible that some of the dimensions are conceptually very similar or are, at least, strongly associated *empirically* (within the domain of emotions investigated). If so, they could be combined for reasons of parsimony. To examine these questions, the statistical associations between the appraisal variables were analyzed. Because of the nominal-scale character of the appraisal scales, we preferred hierarchical cluster analysis for this purpose (e.g., Anderberg, 1973; see also Reisenzein and Hofmann, 1990). The nominal-scale association coefficient lambda (symmetric; see Hays, 1973) was used as the measure of interscale association, and the average linkage algorithm as the method of fusion.[6]

For the study 3 data, the cluster analysis revealed only two clusters of appreciably associated variables. The first cluster comprised subjective evaluation, interpersonal evaluation, and moral evaluation (mean interitem association = .42); the second cluster joined evaluation of other people, social relation positive–negative, and social relation close–distant (.43). The cluster formed at the next step joined self-evaluation and superior–inferior to others, but interscale association was already rather low (.27). For the data from study 4, four clusters were obtained: time-certainty (mean interitem association, .53); evaluation of others, social relation positive–negative, and social relation close–distant (.41); stability–controllability (.40); and causality/responsibility plus self-evaluation (.41). The cluster formed at the next step joined suddenness and expectedness, but again, the association was very low (.22). Hence, only the social evaluation cluster (social relation positive–negative, close–distant, and evaluation of others) was replicated across the two studies.

Reanalyses of the study 3 data for the 22 emotions also used in study 4 replicated the findings for the total data set with the exception that, as in study 4, stability and controllability were now also joined in a cluster (λ = .40). Hence, the differences between the results of the two studies were due only partly to the greater number of emotions included in study 3. A more important reason seems to have been differences in scale use.

[6]Virtually identical solutions were obtained when the analyses were repeated using Cramer's V as the measure of association and Ward's method as the fusion algorithm.

A comparison of the appraisal profiles of the 22 emotions included in both studies for each pair of scales suggested that several scales, in particular certainty, self-evaluation, interpersonal evaluation, familiarity, superior–inferior, close–distant, and stability, were used in a more discriminating manner in study 4. More precisely (a) there was generally higher agreement in study 4 on whether response options *a*, *b*, or c characterized the various emotions, and (b) response option *c* (neither/nor, undecidable, or irrelevant) was used as the most frequent code for a greater number of emotions. Hence, as expected, the "expert" coders — who were familiar with the theoretical definitions of the dimensions — had apparently established sharper meanings of the appraisal dimensions than the subjects, who were presented with the single-item scales without further explanation or training in their use.

The associations between the scales for social relation positive–negative, close–distant, and other evaluation replicated findings by Reisenzein and Hofmann (1990) and may indicate that these scales tap a common, more broadly defined underlying dimension (e.g., liking versus dislike for others). The associations between the scales for subjective, interpersonal, and moral evaluation obtained in study 3 are in accord with comparable findings by Smith and Ellsworth (1985), Frijda *et al.* (1989, study 2), and Manstead and Tetlock (1989). However, they are partly in conflict with the results of study 4, as well as with the results of Reisenzein and Hofmann (1990) and Smith and Ellsworth (1987), who found subjective and moral evaluations to be largely uncorrelated (interpersonal evaluation was not included in these studies). Since these dimensions seem to be clearly distinct conceptually, the associations obtained in study 3 may reflect problems of operationalization; the conflicting findings of other studies can perhaps be attributed to differences in the judged set of emotional situations. The association of stability and controllability replicates findings by Frijda *et al.* (1989) and reflects the empirical phenomenon that stable events tend to be uncontrollable, and vice versa. Again, however, these two dimensions are clearly distinct conceptually (Weiner, 1986) and were not substantially associated in the larger data set including the 30 emotions.

In sum, the majority of the scales was used distinctively and was largely independent within the investigated object domain. The greater number of statistically independent dimensions obtained in the present studies, compared with previous ones, is in part simply due to the fact that more conceptually distinct scales were included to begin with; but it is probably also due partly to the nominal scale rating procedure, which tended to foster dimensional independence (particularly because of the inclusion of an "irrelevant" category), as well as to the method of association analy-

sis.[7] Although a reduction of the 22 scales to some 17 (study 3) or 16 (study 4) may be possible (see also, footnote 7), except for the scales joined in the social evaluation cluster, such a reduction seems to be justifiable only by considerations of parsimony, which, in our view, are of secondary relevance to theory building (cf. Bunge, 1967). For purposes of theory construction, it is of paramount importance to have dimensions that are *conceptually* clear (and distinct), whereas the issue of their empirical independence is secondary (see also, Hofstede, 1980). This is particularly so in view of the fact that the combination of conceptually distinct covarying dimensions can lead to dimensions that are difficult to interpret. Therefore, unless a psychologically meaningful, higher-order dimension can be found under which correlated dimensions can be subsumed, it seems to be most prudent to keep the dimensions separate. This suggestion is reinforced by the finding that even the most strongly associated variables provided independent contributions to emotion prediction (see below).

We would like to emphasize, however, that criteria other than statistical dependence may provide reasons for excluding a candidate from the set of appraisal dimensions (for more detail, see General Discussion). Furthermore, it is of course entirely legitimate to construct, by either combining or excluding dimensions, a *simplified* model of appraisal dimensions that can be more easily handled in practice, provided that one is willing to trade off the increase in parsimony for decreased predictive capacity and inter-

[7]To examine this issue, as well as to enhance the comparability of the results with those of previous investigators, principal-components analyses (PCA) were also computed for the 44 binary dummy variables used in the discriminant analysis reported later [for a discussion of the problems associated with PCA analyses of binary variables, see Comrey (1973)]. The matrix of Pearson correlation coefficients between the dummy variables served as input to PCA; principal components with eigenvalues >1 were retained and rotated according to the varimax criterion. For study 3, 15 factors were obtained, which together explained 74% of the variance of the variables. The first three factors combined precisely the same variables that also marked the first three clusters of the cluster analysis. Factor 4 had high loadings on importance and moderate ones on anticipated effort. The remaining 11 factors corresponded to 11 of the 12 remaining appraisal dimensions; expectedness did not emerge as a separate factor but loaded moderately on several of the other factors (suddenness, predictability, and causality). A nearly identical solution was obtained if only the 22 emotions also used in study 4 were included in the analysis.

For the study 4 data, the PCA resulted in 12 factors that together explained 72% of the variable variance. The results of this analysis were less clearly interpretable. Paralleling the results of the cluster analysis, stability and perceived control were merged in one factor, as were time and certainty/probability. The general, social, and self-evaluation factors of study 3 were now differently represented: (a) A general evaluation factor represented subjective evaluation, positive evaluation of others, and positive and close social relationship; (b) interpersonal and moral evaluation now emerged as a separate factor; and (c) causality/responsibility formed a factor together with negative self-evaluation, negative evaluation of others, and negative social relation. In addition, the intentionality/activity factor also had sizable loadings for self-evaluation; and the focus factor, for positive interpersonal and positive self-evaluation.

pretability (see also, Tesser, 1990). However, it may be more adequate to include in this simplified model those appraisal dimensions that are most salient to the subjects, rather than to base their selection (only) on the results of statistical data reduction techniques.

Discriminatory Capacity of Appraisal Dimensions

As in previous studies (e.g., Frijda *et al.*, 1989; Reisenzein and Hofmann, 1990; Smith and Ellsworth, 1985), discriminant analysis was used to determine how well the scales distinguished between the emotions. For this purpose, the nominal-scale appraisal variables were first transformed into two binary dummy variables each, which served as predictors; the criterion was the nominal scale formed by the emotion categories. Note (a) that the variable transformation is a purely technical device to enable one to use nominal scale variables in the discriminant analysis — the 44 dummy variables carry precisely the same information as do the original 22 ones — and (b) that the dummy variable pairs representing an original variable in the analysis are of course correlated — however, the discriminant procedure takes automatically care of these, as well as any other dependencies among the predictors.

The results of the discriminant analysis are shown in Table III, which also contains, for comparison purposes, discrimination data from Reisenzein and Hofmann (1990, 1993). In study 3 (cf. Table III, column 2), 47.9% of the scenarios for the 30 emotions were correctly classified by the discriminant procedure (κ = .46), ranging from a minimum of 19% (joy) to a maximum of 74% (admiration, relief). In study 4 (cf. Table III, column 4), 71.7% of the scenarios of the 23 emotions were correctly classified (κ = .71), ranging from 40% (hopelessness, remorse, and surprise) to 100% (gratitude). The results were essentially unaltered when stepwise discriminant analysis was used (47.8% correctly classified in study 3 and 70.9% in study 4). Thirty-seven (40) of the 44 dummy variables were retained in the analysis for study 3 (study 4), among them, with the exception of focality–globality in study 4, at least one member of each of the binary variable pairs that represented the original variables in the analysis. This suggests that all variables except focality–globality in study 4 provided an independent contribution to emotion prediction. This finding is actually not surprising, considering (a) that the variables were, as noted previously, largely independent, and (b) that — again, with the exception of focality–globality in study 4 — all variables discriminated significantly between emotions when considered individually (classification accuracy ranged from 5.2 to 9.2% correct in study 3 and from 8 to 13% correct in study 4). The lack

Table III. Percentage of Correctly Classified Situations, Studies 3 and 4

	% correctly classified				
	Study 3		Study 4	Comparison data	
	30 emotions	22 emotions		Reisenzein & Hofmann, 1990 (study 2)	Reisenzein & Hofmann, 1993 (study 1)
Admiration	74	—[a]	—	—	—
Anger/rage	26	48	70	59	74
Anxiety/fear	60	63	90	64	77
Apprehension	44	—	—	—	—
Contempt	41	63	80	82	51
Contentment	56	—	—	—	—
Despair	22	—	—	—	—
Disappointment	26	48	65	64	71
Discontentment with self	48	63	95	50	62
Disgust/revulsion	63	63	80	45	88
Embarrassment	37	41	50	41	55
Envy	48	63	60	91	68
Gloating	50	—	—	—	—
Gratitude	48	67	100	77	77
Guilt	58	54	95	50	55
Hate	39	—	—	—	—
Hope	48	48	75	82	85
Hopelessness/resignation	24	48	40	64	59
Indignation	44	—	—	—	—
Jealousy	41	74	80	41	62
Joy/happiness	19	30	50	77	69
Lonelines	52	52	90	41	76
Longing	44	—	—	—	—
Love	67	78	90	50	52
Pity/Sympathy	67	78	75	86	79
Pride	70	78	95	91	77
Relief	74	67	80	64	81
Remorse/regret	56	52	40	59	32
Sadness/sorrow	44	44	60	64	58
Shame	48	52	50	23	40
Surprise	—	—	40	96	51
Mean	47.9	57.9	71.7	63.5	65.7
SD	15.1	12.9	19.6	19.3	14.5

[a]Emotions not included.

of discrimination of the focality–globality dimension in study 4 was due to the fact that nearly all eliciting events were coded as specific.

The higher percentage of correctly classified situations in study 4 can be attributed to several causes. First, there were fewer emotions and fewer

situations per emotion than in study 4. If only the 22 emotions also examined in study 4 were considered in the discriminant analysis for the study 3 data, the average percentage of correctly classified situations increased to 57.9% (κ = .56; cf. Table III, column 3); and if additionally only 20 (randomly selected) situations per category were included (as in study 4), discrimination increased further to 62.1%. Second, as mentioned before, several scales were used in a more sharply discriminating manner in study 4. Third, it is likely that the aspect of the situation at which the emotion and its correlated appraisals were directed was more clearly specified in the scenarios (cf. Reisenzein and Hofmann, 1993). To illustrate, whereas in study 4 none of the pride situations but 80% of the anxiety and 70% of the hope situations were coded as uncertain, in study 3 fully 52% of the pride situations but only 59% of the anxiety and hope situations were judged as uncertain. Quite possibly, the "uncertain" judgments for pride and the "certain" judgments for anxiety and hope referred to an aspect of the eliciting event different from the object of the emotions (which may also have been encouraged by the specific wording of the certainty item; cf. the Appendix).

The correlation of the percentages of correctly classified scenarios obtained in studies 3 and 4 (for the 22 emotions considered in both studies) was .63, and the average correlations between the classification results of study 3 and 4 and those of the comparison studies (cf. Table III, columns 5 and 6) were only .18 and .25. Hence, different emotions were well vs. badly discriminated in the different studies. In part, this can be attributed to procedural differences. Reisenzein and Hofmann (1990) used a somewhat different set of predictors and the subjects indicated the appraisals for situations *typically* conducive to the various emotions, whereas the results of Reisenzein and Hofmann (1993) are based on subjects' classifications of the 460 scenarios used in study 4. In any case, the different rank orders of discrimination accuracy obtained in the various studies suggest that the limit of discrimination of emotions by cognitive appraisals has not yet been reached.

Typicality and Generality of Appraisal Components

As a further index of the importance of a proposed dimension of appraisal, we computed, for each of the 22 scales, the maximum percentage of endorsements of their substantive dimension values (either *a* or *b*, whichever was larger) for the different emotions. This index can be regarded as a measure of the perceived *typicality* or centrality of the appraisal component *a* or *b* for an emotion (Smith & Medin, 1981). In general, keeping

other factors constant, an appraisal component — and, by extension, the dimension to which this component belongs — can be regarded (a) as more important for a given emotion, the higher its typicality for that emotion, and (b) as more important for a set of emotions, the greater the number of emotions from this set for which it has high typicalities or, as we say, the greater its *generality* across this set.

The study 3 results of these analyses are contained in Table IV, which shows, in simplified form, the appraisal patterns for the emotions obtained in this study.[8] As can be seen, all 22 appraisal dimensions had typicalities (percentage endorsements) ≥60%, and all except suddenness, expectedness, certainty, predictability, and stability had typicalities ≥80% *for at least one emotion*, although these high endorsements sometimes concerned only one of the substantive dimension values *a* or *b*. In study 4, all dimensions except moral evaluation, superior–inferior, and familiarity had typicalities ≥80% for at least one emotion. Across both studies, the lowest maximum typicality value for any emotion was obtained for the dimension value *global* (30% in study 3). The next-lowest maximum typicalities were found for *expected* (70%; study 3) *superior*, *novel*, *unpredictable*, and *low anticipated effort* (maximum, 75%).

There were, however, substantial differences with regard to the dimensions' *generality*, operationalized as the number of emotions for which high typicalities (endorsements ≥80%) on at least one of the dimension values were obtained. For example, whereas subjective evaluation had high typicalities for 24 (80%) of the 30 emotions in study 3 (cf. Table IV), and for 21 (91%) of the 23 emotions in study 4, predictability of consequences had high typicalities for only 1 emotion (hopelessness) in study 4. This finding confirms the assumption of most cognitive emotion theorists that although a few appraisal dimensions are important for most emotions, many are relevant only to a subset. The rank-order correlation between the generality scores of the dimensions (cf. Table IV for those of study 3) and the frequencies with which they were coded in the grid studies (cf. Table II) was .58 ($p < .01$) for study 3 and .53 ($p < .05$) for study 4. Hence, the more general a dimension, the more likely it tended to be spontaneously mentioned.

Summary. (1) The majority of the appraisal scales was statistically independent, indicating that they not only are different in meaning, but also are largely uncorrelated within the investigated domain of emotion-eliciting situations. (2) The 22 scales contributed significantly to emotion discrimination both when considered jointly and separately. (3) A higher percent-

[8]The exact appraisal patterns for study 3 as well as study 4 can be obtained from the first author.

Table IV. Study 3: Appraisal Patterns[a]

Emotion	Subjective evaluation + −	Interpersonal evaluation + −	Moral evaluation + −	Self evaluation + −	Other evaluation + −	Social relation + −	Superior–inferior	Close–distant	Important–unimportant	Present–future	Sudden–gradual	Expected–unexpected	Familiar–unfamiliar	Uncertain–certain	Predictable–unpredictable	Unstable–stable	Controllable–uncontrollable	Responsibility self–other/cir.	Intentionality self/other–unint.	Focus self–other	Anticipated effort high–low	Specific–global
	Appraisal dimension																					
Hopelessness	− −	−	∘	−	∘	∘	∘	−	+ +	+ +	∘	∘	+	+	−	∘	∘	−	∘	∘	∘	+ +
Despair	− −	−	∘	∘	∘	∘	−	−	+ +	+	∘	∘	∘	+	∘	∘	+	−	∘	+	+ +	+
Loneliness	− −	−	∘	∘	∘	∘	−	−	+	+ +	∘	∘	∘	+	∘	∘	+	−	∘	+ +	∘	+
Anxiety	− −	∘	∘	∘	∘	∘	−	−	+	∘	∘	−	−	∘	∘	−	+ +	∘	∘	+ +	∘	+ +
Sadness	− −	−	∘	∘	∘	∘	∘	∘	+ +	+ +	∘	∘	∘	∘	∘	∘	∘	∘	∘	∘	+	+ +
Remorse	−	−	−	− −	−	∘	∘	∘	+	+ +	∘	∘	+	∘	∘	∘	+	+ +	∘	∘	∘	+ +
Guilt	− −	−	∘	− −	−	∘	∘	∘	+	+ +	∘	∘	∘	∘	∘	+	+	+ +	∘	∘	+	+ +
Shame	− −	−	∘	− −	−	∘	−	∘	∘	+ +	∘	∘	∘	∘	∘	∘	∘	+ +	∘	+	∘	+ +
Discontentment	− −	−	∘	− −	∘	∘	− −	∘	+	+ +	−	∘	+	∘	∘	+	+	+ +	∘	∘	+	+
Embarrassment	−	∘	∘	−	∘	∘	−	∘	∘	+	∘	∘	∘	∘	∘	∘	∘	∘	∘	∘	−	+
Indignation	− −	−	−	+	−	−	∘	−	+	+ +	+	−	∘	∘	∘	∘	∘	−	∘	+	+	+ +
Hate	− −	− −	− −	∘	−	−	∘	−	+	+ +	∘	−	∘	∘	∘	∘	∘	−	∘	∘	+	+ +
Anger	− −	−	−	∘	−	−	∘	−	+	+ +	∘	−	+	∘	∘	∘	∘	−	∘	+	∘	+ +
Contempt	− −	−	−	∘	−	−	+	− −	∘	+ +	∘	−	∘	∘	∘	∘	∘	−	∘	∘	∘	+ +
Jealousy	− −	∘	−	∘	−	−	− −	−	∘	+ +	+	∘	+	∘	∘	∘	∘	−	∘	+	+	+ +
Disappointment	− −	−	∘	∘	−	−	∘	−	+	+ +	∘	−	∘	∘	∘	∘	∘	−	∘	∘	∘	+ +
Disgust	− −	−	∘	∘	−	∘	∘	∘	∘	+ +	∘	−	∘	∘	∘	∘	∘	−	∘	∘	∘	+ +
Gloating	−	− −	∘	∘	∘	−	+	∘	∘	+ +	∘	∘	∘	+	∘	∘	∘	−	∘	− −	∘	+ +
Envy	−	∘	∘	−	∘	−	− −	−	∘	+ +	∘	∘	∘	∘	∘	∘	∘	−	∘	∘	∘	+

Table IV. Continued

Emotion	Appraisal dimension: Subjective evaluation + −	Interpersonal evaluation + −	Moral evaluation + −	Self evaluation + −	Other evaluation + −	Social relation + −	Superior_inferior	Close_distant	Important_unimportant	Present_future	Sudden_gradual	Expected_unexpected	Familiar_unfamiliar	Uncertain_certain	Predictable_unpredictable	Unstable_stable	Controllable_uncontrollable	Responsibility self_other/cir.	Intentionality self/other_unint.	Focus self_other	Anticipated effort high_low	Specific_global
Apprehension	− −	○	○	○	○	+	○	+	++	+	○	!	○	○	○	!	+	○	○	− −	○	++
Pity	−	○	○	○	○	++	○	+	+	++	○	○	○	○	−	!	○	−	○	− −	○	++
Pride	++	++	++	++	○	○	+	○	++	++	−	○	○	○	○	○	○	++	++	++	○	++
Contentment	++	++	+	++	○	○	○	○	++	++	○	○	+	○	○	+	+	+	+	++	○	+
Relief	++	++	+	○	○	○	○	○	++	++	○	+	○	○	+	○	○	○	+	++	○	++
Longing	+	○	○	○	+	+	○	○	+	+	○	○	++	○	○	○	○	○	○	+	○	+
Hope	++	+	○	○	+	+	○	○	++	○	○	○	○	○	○	○	++	○	+	+	○	+
Gratitude	++	++	+	+	++	++	○	++	++	++	○	○	○	○	○	○	○	−	○	○	○	++
Joy	++	++	+	+	+	+	○	++	++	++	○	○	○	○	○	○	○	○	○	○	○	++
Love	++	++	+	+	++	++	○	++	++	++	−	○	+	○	○	○	+	○	○	○	○	+
Admiration	++	++	+	○	++	++	○	○	○	++	○	○	○	○	○	○	○	− −	○	−	○	++
Generality index[b]	80	30	07	20	23	30	10	20	37	80	00	00	03	00	00	00	07	33	03	33	03	67

[a]Symbols represent the percentage of endorsement of the A and B responses as follows: ++, % of A responses ≥80%; +, % of A responses ≥60% but <80%; − −, % of B responses ≥80%; − % of B responses ≥60% but <80%.

[b]Percentage of emotions with endorsements of A or B response ≥80% (++, − −)

age of situations was correctly classified in the expert coding study. This seemed to be due partly to a more discriminate use of the scales, suggesting that prediction might be improved if the appraisal dimensions were more adequately operationalized or subjects were trained in the use of the scales. (4) The dimensions which emerged as the most salient ones in the grid studies had high typicality values for particular emotions and tended to be those relevant for the greatest number of emotions.

GENERAL DISCUSSION

Taken together, the results of the studies reported in this article can be summarized as follows:

1. The majority of the examined candidates for appraisal dimensions seems indeed to refer to descriptive or evaluative features of the situation that are spontaneously considered by people, at least when they reflect on such differences after the event (see below for an elaboration of this point).
2. The dimensions regarded as most important in previous theory and research agree only partly with those that are seen as the most salient ones by people, whereas other dimensions that have been comparatively less featured in previous research (such as focus, self- and other evaluation, and dimensions pertaining to social relationship aspects) are relatively salient. It is interesting to note that the latter dimensions seem to be particularly relevant for interpersonal emotions; hence, Markus and Kitayama's (1991) suggestion that such appraisal dimensions (specifically, focus) may be unique to, or at least particularly salient in, non-Western cultures finds little support in our findings.
3. The list of appraisal dimensions seems to be fairly complete, at least for the emotions that were included.
4. The 22 dimensions examined in studies 3 and 4 seem to be for the most part conceptually distinct and statistically independent within the investigated object domain.
5. These 22 dimensions also provide independent contributions to the statistical discrimination among emotions, indicating that they are potentially suited to serve as the cognitive discriminantia of emotions in the process of appraisal, as well as in inferences of emotions from situational information (cf. Reisenzein and Hofmann, 1993).
6. Most of the 22 dimensions also have at least one substantive value that is from fairly to highly typical for at least one of the included

emotions; but they differ pronouncedly in their generality, i.e., the number of emotions for which they are seen as highly relevant. Furthermore, the greater the generality of a dimension, the more likely it was mentioned spontaneously in the grid studies.

In terms of the goals of this research — to help demarcate the set of dimensions of emotional appraisal in emotion — then, the results suggest that the original list of candidates of appraisal dimensions was, by and large, well chosen. More precisely, the results give no reason for enlarging the initial set of candidates, and they suggest that this set can be somewhat, but not drastically reduced by excluding or combining dimensions.

However, the results are certainly not strong enough to warrant the conclusion (a) that there are no further dimensions of appraisal or (b) that all of the candidates of appraisal dimensions that "passed" the combined tests to which they were subjected in the present studies must therefore necessarily be accepted as being dimensions of *appraisal.* These conclusions would be unwarranted, first, because of inevitable methodological limitations of the present research and, second, because the criteria used in the present studies for evaluating the status of a proposed dimension of appraisal may be insufficient, too weak, or even partly inadequate. The remainder of the General Discussion is devoted to a more thorough examination of these issues.

Are There Further Dimensions of Appraisal?

Although the results of the grid studies provided no evidence for further dimensions of appraisals, the conclusions that can be drawn from this finding are necessarily restricted by the limited samples of emotions and comparisons between their eliciting situations. However, there is some reason for believing that these limitations were not too serious. First, the emotions constituted a fairly representative sample of the affective spectrum (perhaps somewhat biased toward the interpersonal side). Second, there were no significant differences between the response distributions for differently sampled subgroups of comparisons in study 1. Nevertheless, given that the appraisal dimensions differed considerably in generality, it is possible that further dimensions of appraisal — that are relevant only to specific emotions — might emerge if different emotions were studied or subsets of highly similar emotions were compared (see, e.g., Smolenaars & Schutzelaars, 1986/1987). Also, further or different dimensions might emerge in different cultures. Future studies of appraisals using the grid technique should therefore particularly focus on these issues.

In addition, there are more general methodological and theoretical considerations that could lead one to expect additional dimensions of appraisal even for the investigated emotions. First, it could be argued that the grid method is not suited to elicit all relevant appraisal dimensions; in particular, (a) it may be unsuited to detect subtle cognitive distinctions between emotions or (b) some appraisals may be in principle unconscious and therefore not accessible by this method.

To the first of these concerns, we reply that, inasmuch as many rather subtle cognitive distinctions *did* emerge in the grid studies, those that failed to do so are probably only of very subordinate importance. That is, they are — as supported by the correlation between the frequencies of the dimensions and their generality across emotions — relevant only to very few emotions or are even restricted to a small subgroup of the situations conducive to a particular emotion. To the second concern, we reply that a distinction must be drawn between the *process* of appraisal and the *results* or outcomes of this process (see also Lazarus & Smith, 1988). Although the process of appraisal may typically be unconscious, we assume that its outcomes are normally consciously accessible; and the present studies were concerned only with the outcomes of the appraisal process. This assumption is in fact shared by most appraisal theorists. In particular, all of the appraisals examined in the present studies are assumed to be consciously accessible and measurable by self-report by the authors who proposed them. Admittedly, these appraisals may be denied, distorted, or repressed by some people some of the time. However, this is certainly not always the case, and there is little reason for believing that such factors are operative if abstract emotion-eliciting situations are compared (study 2).

Hence, although the existence of kinds of appraisal that are not consciously accessible in principle is a logical possibility, these appraisals are certainly very different in nature from those examined in the present studies. Appraisals that are unconscious in principle cannot contribute directly to the distinctions commonly made between emotions, which is often regarded as one of the major explanatory functions of appraisals (cf. Reisenzein & Schönpflug, 1992); in addition, such appraisals cannot be detected or validated by *any* method that relies on verbal reports. Because all known methods used to measure appraisals ultimately rely on self-reports, this means that the existence of such appraisals is currently not demonstrable at all, except perhaps in rather indirect ways.

A second consideration that could be advanced in support of the existence of further dimensions of appraisal even for the emotions studied is the following one. As mentioned in the Introduction, appraisal theorists assume typically — although not inevitably — that each distinct emotion is characterized by a distinct pattern of appraisal (cf. Reisenzein and Hof-

mann, 1993). If so, perfect discrimination between emotions on the basis of appraisals should theoretically be possible. The statistical discrimination accuracies obtained in the present studies still fall considerably short of this theoretical expectation, although they begin to approximate subjects' classification of emotions on the basis of appraisal-relevant situational information (cf. Reisenzein and Hofmann, 1993). Therefore, it may be concluded, further appraisal dimensions are necessary. However, as discussed in more detail by Reisenzein and Hofmann (1993), the moderate degree of appraisal discrimination of emotions obtained was probably due to a fair degree to methodological problems. These include the possibility (a) that some of the emotions studied were not regarded as distinct by some people to begin with, (b) that some of the appraisal scales suffered from low reliability or validity, (c) that the objects of the various appraisals were not identified with sufficient precision [which is particularly important if a situation evokes multiple emotions (cf. Reisenzein and Hofmann, 1990; Scherer, 1993)], and (d) that the statistical prediction models used were not fully adequate for the task (cf. Gehm & Scherer, 1988; Tesser, 1990). These factors were undoubtedly also operative to varying degrees in the present studies 3 and 4 (cf. the previous discussion of these studies) and may together be sufficient to explain the less than perfect discrimination between emotions that was obtained.

Are All of the Substantiated Dimensions Dimensions of Appraisal?

To achieve its intended goal, research on the dimensions of emotional appraisal presupposes clarity about the *criteria* by which appraisal dimensions are to be identified and distinguished from nonappraisal dimensions. These criteria include both conceptual–theoretical and empirical ones. The former concern the question of what one is to understand, conceptually, by an appraisal dimension, whereas the latter concern the question of, once this has been determined, how one can empirically identify the appraisal dimensions actually used by subjects. Several such criteria have been implicitly used in the appraisal research existing to date. Rarely, however, has the issue of the theoretical foundation and adequacy of these criteria, as well as the question of further possible criteria for appraisal dimensions, been explicitly discussed (cf., in particular, Lazarus & Smith, 1988; Reisenzein and Hofmann, 1990, 1993). The possibility therefore exists that the criteria used in previous, as well as in the present studies, to demarcate the set of appraisal dimensions were insufficient, too weak, or even partly inadequate. If so, different conclusions would be reached if additional or different criteria were used. A thorough review and discussion of this issue

are much beyond the goals of the present article, but we reflect at least briefly on those criteria that were used in the present studies as well as on further potential criteria for appraisal dimensions.

In the present studies, five criteria for dimensions of appraisal in emotions were used. We started out with a theoretically derived list of candidates for appraisal dimensions. The members of this list already had to fulfill what we believe to be a minimal necessary conceptual criterion of appraisal dimensions, namely, (1) that they can be taken to refer to a possible cognitive or evaluative judgment of (an aspect of) the eliciting situation. The results of the grid studies constituted (2) the first empirical indicator (spontaneous concept usage) of the validity of this list, and studies 3 and 4 provided three more empirical criteria concerning, respectively, (3) the redundancy vs. independence of the dimensions, (4) the capacity of the potential appraisal dimensions to discriminate, singularly and in combination, between the emotions (see also Reisenzein and Hofmann, 1993), and (5) the typicality of the values of the dimensions for different emotions.

Fulfillment of criteria 1, 4, and 5 may be regarded as necessary, at least if they are appropriately specified. That is, no dimension should be regarded as a dimension of appraisal unless it refers to a cognitive or evaluative aspect of the eliciting (appraised) situation, discriminates to some degree between emotions (or at least between emotions and nonemotions), and unless at least one of its values is typical for at least one emotion. Fulfillment of criterion 2 (the dimension shows up in Grid studies) is not necessary, but we think that any dimension that does not pass this test even after repeated attempts should be regarded with some doubt. Finally, fulfillment of criterion 3 (statistical independence) is certainly not necessary, although complete statistical dependence of a dimension on others in the universe of emotion-eliciting situations would be a reason for excluding that dimension. Furthermore, the more of these conditions are fulfilled by a candidate appraisal dimension, the more confident we can be that it truly is a dimension of *appraisal.*

Beyond the criteria used in the present studies to delineate the set of appraisal dimensions, two further criteria may be useful. The first of these was suggested by Reisenzein and Hofmann (1990): A proposed appraisal dimension should be accepted as a genuine dimension of appraisal only if it can be plausibly regarded as a judgment or evaluation that the person *actually makes* (if only implicitly) during the process of interpreting emotion-eliciting events, in contrast to a judgment that occurs only post-hoc, when one reflects on differences between emotion-eliciting situations. This criterion may not be fulfilled, for example, by the dimension of *focality-globality* (Frijda, 1986). Although some emotion-eliciting events are unquestionably more diffuse in character than others, we doubt that people, during the process of appraisal,

judge an event as being specific versus unspecific and, depending on the outcome of this judgment, experience different emotions. It seems more plausible to us to assume that this dimension of emotion-eliciting situations is the result of a post hoc reflection on differences between eliciting events.

The second criterion, which constitutes a radicalization of the typicality criterion mentioned above, is that only those dimensions should be regarded as dimensions of appraisals whose values belong to the necessary or defining features of at least one emotion; hence, if that appraisal component is not present, one cannot say that the emotion is present either. This criterion is implicit in the cognitive emotion theories proposed, for example, by Stumpf (1899; cf. Reisenzein & Schönpflug, 1992), Johnson-Laird and Oatley (1989; for "complex emotions"), and Ortony *et al.* (1988). Theoretically, this criterion would permit a sharp delineation of appraisal dimensions. However, it has two problems: First, it presupposes that emotions have at all necessary or defining features, which is denied by several authors (see e.g., Russell, 1991); and second, it is no easy task to demonstrate empirically the presumed indispensability of an appraisal component for an emotion. Nevertheless, further elucidation of this criterion seems to be a promising line of future inquiry. In any case, we believe that future research in this area would profit much from explicit discussions of criteria for dimensions of appraisal in emotion.

APPENDIX

Coding System for Appraisal Dimensions

For each of the following *potential* dimensions of appraisal, a brief theoretical characterization is given and important literature sources (mostly restricted to the recent appraisal literature) are cited. In addition, for the 22 dimensions included in studies 3 and 4, the questionnaire items used to measure the dimensions, as well as examples of words or phrases coded as referring to the dimensions in studies 1 and 2, are listed.

Evaluative Dimensions

1. Subjective Evaluation of the Eliciting State of Affairs (ESA). Evaluation of ESA on a dimension ranging from desirable/positive to undesirable/negative for self. The standards of evaluation are purely personal or subjective (in contrast to interpersonal or moral; cf. categories 3 and 4), such as momentary or long-standing personal goals, preferences and de-

sires. (This category was also coded if the standard of evaluation could not be determined.)

Literature: This or a similar dimension is ubiquitous, e.g., evaluation based on personal criteria (Solomon, 1976), valence of outcome (Weiner, 1982), or desirability (Ortony *et al.*, 1988). Scherer (1984, 1988) and Smith and Ellsworth (1985) distinguish between the intrinsic pleasantness of an event and its goal conduciveness (category 2). See also motive consistency (Roseman, 1984) and goal congruence versus incongruence (Lazarus, 1991).

Examples: Situation positive–negative, desirable–undesirable, something good–bad happens to me, event is agreeable–disagreeable, success–failure, gain–loss, problem solution–no solution.

Questionnaire Item: In this situation, something happens which is desirable/positive–undesirable/negative for me.

2. Goal Conduciveness. The degree to which ESA is perceived as furthering versus hindering one's goals or plans (includes perceived obstacles or barriers). Appraisals on this dimension may be regarded as underlying a subset of subjective evaluations (category 1).

Literature: Blocking of goal attainment (Krech & Crutchfield, 1958; Hunt *et al.*, 1958), goal conduciveness (Scherer, 1988; Gehm & Scherer, 1988), and goal-path obstacle (Smith and Ellsworth, 1985; compare also Frijda *et al.*, 1989).

3. Interpersonal Evaluation of ESA. The emotion experiencer is concerned about the evaluation of ESA by significant others, about whether others' expectations have been met or not met, etc.

Literature: Evaluation based on interpersonal criteria (Solomon, 1976); internalized social standards (Krech & Crutchfield, 1958), compatibility with external standards (Scherer, 1988; Gehm & Scherer, 1988), and consistency with other's standards (Manstead & Tetlock, 1989); compare also type of ego involvement — social esteem (Lazarus, 1991).

Examples: Something that others regard as good-bad, others' expectations fulfilled–not fulfilled, my work, competence, achievement is acknowledged–not acknowledged by others, other people apply standards–do not apply standards, approval–disapproval by others.

Questionnaire Item: In this situation, something happens which others would evaluate positively/approve of–which others would evaluate negatively/disapprove of.

4. Moral Evaluation of ESA. ESA (typically an own action or performance) is evaluated by the emotion experiencer as morally or ethically good versus bad in some respect. In contrast to subjective and interpersonal evaluation, the standards of moral evaluations are typically regarded as objectively valid.

Literature: Evaluations based on moral criteria (Solomon, 1976; see also Krech & Crutchfield, 1958), legitimacy/fairness (Roseman, 1979; Smith and Ellsworth, 1985), norm compatibility (Scherer, 1988), and type of ego-involvement — moral values (Lazarus, 1991).

Examples: Morally good–bad, fair–unfair, deserved–undeserved, justice–injustice, honest–dishonest act, I/someone else did something right–wrong, violates my system of values–does not violate my value system.

Questionnaire Item: In this situation, something happens which is right, just, fair, or deserved–wrong, unjust, unfair, or undeserved.

5. Self-Evaluation. The emotion experiencer evaluates him- or herself positively or negatively in some respect (frequently as a consequence of his or her evaluation of an ESA).

Literature: A subtype of Frijda's (1986) and Ortony and co-workers' (1988) object evaluation; compatibility with self-image (Scherer, 1988; Gehm & Scherer, 1988), self-esteem (Frijda *et al.*, 1989), type of ego involvement — self-esteem (Lazarus, 1991); similar to consistency with personal standards (Manstead & Tetlock, 1989).

Examples: I am great–not great, I am important–unimportant, I am valuable–worthless, I am competent–incompetent, I respect–do not respect myself, I like myself–dislike myself.

Questionnaire Item: In this situation, I evaluate myself positively–negatively.

6. Evaluation of Others. The emotion experiencer evaluates another person involved in ESA positively or negatively in some respect.

Literature: Subtype of object evaluation (Frijda, 1986; Ortony *et al.*, 1988).

Examples: Somebody else is great–not great, competent–incompetent, valuable–worthless, I respect–do not respect other, I take someone seriously–do not take him or her seriously, I like–dislike somebody else.

Questionnaire Item: In this situation, I evaluate another person positively–negatively.

7. Evaluation of Social Relation (Appraisal of Social Relationship I). This category comprises positive or negative evaluations, in some respect, of the relationship that momentarily exists to another person (other people) involved in ESA (momentary evaluations rather than longstanding evaluative dispositions are meant).

Literature: Suggested by data from Reisenzein and Hofmann (1990), but compare also Solomon's (1976) judgments of "intersubjectivity," Mees (1985), and Kemper (1978).

Examples: Positive–negative relationship, I feel attracted–repulsed by other, trust–no trust, accepted–not accepted by other, understanding–no understanding for other, understanding–no understanding from other.

Questionnaire Item: In this situation, I have a positive–negative relationship to another person.

8. Superiority–Inferiority (Appraisal of Social Relationship II). This category captures social comparison-based judgments of the form that one perceives oneself to be in some respect superior or inferior to other people involved in ESA; it could be regarded as a combination of a special form of self- and other-evaluation. Again momentary evaluations rather than dispositional attitudes are meant.

Literature: Regarded as a basic appraisal dimension of social emotions by Krech & Crutchfield (1958); compare also Solomon's (1976) "personal status."

Examples: I am weaker–stronger than other, other is better–worse than I, other is superior–inferior to me, I am on same level as other–I look down at other.

Questionnaire Item: In this situation, I feel superior–inferior to other(s).

9. Closeness to–*Distance from Other(s) (Appraisal of Social Relationship III).* Judgment that one feels physically or psychologically close to vs. distant or separated from others involved in ESA. Again, momentary judgments rather than dispositional attitudes are meant.

Literature: Suggested by data from Reisenzein and Hofmann (1990), but compare also Solomon's (1976) "intersubjectivity" judgments and Kemper (1978).

Examples: Distanced from–close to others, open to others–isolated from others, I am on my own–together with others, group–alone.

Questionnaire Item: In this situation, I feel close to–distant/separated from other(s).

10. Importance of ESA. Appraisal of the importance or relevance of ESA, e.g., with respect to the centrality of goals or values that are affected, the number and temporal duration of consequences, or the urgency of necessary actions.

Literature: Importance (Brown & Weiner, 1984), concern relevance/urgency (Scherer, 1988), relevance, seriousness, urgency (Frijda, 1986), importance (Frijda et al., 1989), goal relevance (Lazarus, 1991).

Examples: ESA is important–unimportant, has important effects–no important effects, central–peripheral, existential event–not existential, affects me deeply–superficially, I do not care–I hope for change, I am totally involved–I am indifferent.

Questionnaire Item: In this situation, something important–unimportant happens.

Nonevaluative Dimensions

11. Time of ESA. ESA is perceived as past, present, or future.

Literature: Time of event (Scherer, 1984; Frijda *et al.*, 1989); see also Hunt, Cole, and Reis (1958).

Examples: ESA present–future, still present–already past, emotion concerns a future–past event, event occurred–prospect of event, future-related–related to past, emotion occurs during–after an event.

Questionnaire Item: This situation is concerned with something that is present/has already happened–something future/something which has not yet happened.

12. Suddenness of ESA. ESA occurred quickly (including the case that a goal was quickly reached) versus developed slowly (a goal was slowly reached).

Literature: Krech & Crutchfield (1958), Frijda (1987; Frijda *et al.*, 1989), and Scherer (1988).

Examples: ESA occurs suddenly–occurs gradually, sudden change–gradual development, slow–quick change, spontaneous occurrence–ESA has a long history.

Questionnaire Item: In this situation, something occurs suddenly–this is not the case.

13. Expectedness of ESA. Concerns the question of whether the person had an ESA-related expectation or not and, if such an expectation was present, whether it was confirmed vs. disconfirmed by the event (unexpected events include those concerning which the person had no specific event-related expectation at all).

Literature: Unexpectedness (Ortony *et al.*, 1988; Frijda *et al.*, 1989), predictability, expectation (Scherer 1988; Gehm & Scherer, 1988; see also Abelson, 1981; quoted by Mandler, 1984).

Examples: ESA is something I expected to happen–something unexpected, surprising–unsurprising, expectation confirmed–disconfirmed.

Questionnaire Item: In this situation, something happens which I expected–something happens which I did not expect/was unexpected.

14. Familiarity of ESA. ESA is something with which the person is familiar versus something that is novel, unfamiliar or strange.

Literature: strangeness–familiarity (Frijda, 1986), novelty (Scherer, 1988), familiarity (Frijda *et al.*, 1989).

Examples: Familiar–unfamiliar/novel event, old–new, something I know–I don't know.

Questionnaire Item: In this situation, something happens which is familiar/known–unfamiliar/novel to me.

15. Certainty/Probability of ESA. Concerns the perceived certainty or subjective probability of ESA.

Literature: Probability (Roseman, 1979; 1984), expectancy (Pekrun, 1984), certainty (Smith and Ellsworth, 1985; Frijda *et al.*, 1989), likelihood (Ortony *et al.*, 1988), outcome probability (Scherer, 1988).

Examples: Uncertain how event will end–certain, situation is uncertain–certain, event is improbable–probable, event is still uncertain–event has occurred, I know–do not know how things will turn out.

Questionnaire Item: This situation concerns something of which I am uncertain whether it will happened or has happened–something of which I am certain that it has happen or will happen.

16. Predictability of ESA Consequences. Degree to which the consequences of ESA are perceived as predictable vs. unpredictable in various respects by the emotion experiencer.

Literature: Clearness (Frijda *et al.*, 1989) and certainty of consequences (Roseman *et al.*, 1990); compare also uncertainty (Smith and Ellsworth, 1985).

Examples: Consequences/reaction calculable–not calculable, consequences predictable–unpredictable, consequences clear–unclear, situation clear–diffuse, concerns something that is settled–something that is not yet settled, I know–do not know how others will be affected.

Questionnaire Item: In this situation, something happens whose consequences I can foresee/predict–cannot foresee/predict.

17. Stability of ESA. Degree to which ESA is perceived as something stable, lasting or permanent vs. something short-lived, momentary, or changing.

Literature: Modifiability-finality (Frijda, 1986; Frijda *et al.* 1989), stability (Weiner, 1982), future expectancy (Lazarus, 1991).

Examples: Event is final–not final, permanent–short-lived, will change again–will not change any more, temporary–permanent change, no end in sight–closed issue, unique–lasting situation.

Questionnaire Item: In this situation, something happens which may change again (to the better or the worse)–which is unlikely going to change again.

18. Controllability of ESA. The event (or, if the event has already occurred, its consequences) is viewed as controllable vs. uncontrollable by the emotion experiencer.

Literature: Power (Solomon, 1976; Roseman, 1984; Roseman *et al.*, 1990), self-control (Smith and Ellsworth, 1985), control/power (Scherer, 1984, 1988; Gehm & Scherer, 1988; Lazarus, 1991), controllability (Weiner, 1982).

Examples: I can change/influence ESA–cannot change/influence ESA, I am powerful–helpless, something can still be done–nothing can be done any more, I can prevent event–cannot prevent event, I can make up for ESA–cannot make up for ESA.

Questionnaire Item: In this situation, something happens which I can (still, or again) change or influence–something happens which I cannot (or can no longer) change or influence.

19. Causality/Agency/Responsibility for ESA. Subsumes judgments concerning the perceived cause or agent of ESA (self, other, impersonal circumstances) and judgments of responsibility (who is held responsible for ESA).

Literature: Agency (Roseman, 1979), locus of causality (Weiner, 1982), responsibility (Smith and Ellsworth, 1985, Solomon, 1976), agency (Frijda *et al.*, 1989; Scherer, 1988), attribution (Ortony *et al.*, 1988); compare also accountability/blame and credit (Lazarus, 1991).

Examples: Event caused by self–by other, event dependent on me–on external circumstances, I–both are causes, I am responsible–others are responsible, other is responsible–nobody is responsible, my fault–other's (or nobody's) fault.

Questionnaire Item: In this situation, something happens which was caused primarily by me/for which primarily I am responsible–which was caused primarily by somebody or something else/for which primarily somebody or something else is responsible.

20. Intentionality/Activity Concerning ESA. Judgment concerning whether ESA was produced intentionally by oneself or whether one was actively involved in bringing about or preventing an event (or in an attempt to do so) versus not (i.e., somebody else or nobody intended to bring about or to prevent ESA).

Literature: Intentionality (Frijda, 1986; Weiner, 1986), activity (Reisenzein and Hofmann, 1990).

Examples: I am actively involved–I am passive, I do something positive (negative)–I am a bystander, willed–not willed, deliberately caused–accident.

Questionnaire Item: In this situation, something happens which I actively tried (or try) to bring about or to prevent–which I did not (or do not) actively try to bring about or to prevent.

21. Focus of ESA. Concerns the issue of who is primarily affected by ESA (oneself vs. somebody else).

Literature: Focus (Weiner, 1982; Ortony *et al.*, 1988), object fate vs. subject fate (Frijda, 1986), someone else (Frijda *et al.*, 1989); compare also type of ego-involvement — other persons and their well-being (Lazarus, 1991).

Examples: Concerns me–concerns somebody else, my problem–problem of other person, affects me directly–only indirectly, I am directly affected–I am bystander.

Questionnaire Item: In this situation, something happens which concerns primarily myself–which concerns primarily somebody else.

22. Anticipated Effort. Extent to which the person perceives the necessity to undertake physical or mental effort to deal with a present or future event or its consequences.

Literature: Smith and Ellsworth (1985, 1987), Frijda *et al.* (1989); compare also "effort calculation" (Kukla, 1972; Meyer, 1973) and Scherer's (1988) adjustment check.

Examples: I must exert effort–I can relax, something must be done–nothing needs to be done, I am under strain–there is no strain, event has been mentally worked out–has not yet been worked out, event poses further demands–no further demands.

Questionnaire Item: I will have to exert much effort to deal with this situation–I won't have to exert much effort to deal with this situation.

23. Focality–Globality. Degree to which ESA is perceived as something concrete and specific vs. something unspecific or diffuse.

Literature: Focality–globality (Frijda, 1986; 1987; Frijda *et al.*, 1989); compare also Solomon's (1976) scope and focus.

Examples: Event is clear–vague, event is definable–undefinable, elicitor obvious–not obvious, specific–diffuse eliciting event.

Questionnaire Item: In this situation, my emotion is elicited by a concrete event–by unspecific things.

24. Difficulty. Perceived difficulty of a situation or task.

Literature: Smith and Ellsworth (1987).

25. Interestingness. Judgment of how interesting the event or situation is.

Literature: Frijda (1986; Frijda *et al.*, 1989). According to Frijda (1987), this dimension is also related to Scherer's (1984) novelty and Smith and Ellsworth's (1985) attentional activity.

REFERENCES

Adams-Webber, J. (1979). *Personal construct theory: Concepts and applications.* New York: Wiley.

Anderberg, M. R. (1973). *Cluster analysis for applications.* New York: Academic.

Arnold, M. B. (1970). Perennial problems in the field of emotion. In M. B. Arnold (Ed.), *Feelings and emotions* (pp. 169-185). New York: Academic Press.

Brown, J., & Weiner, B. (1984). Affective consequences of ability versus effort ascriptions: Controversies, resolutions, and quandaries. *Journal of Educational Psychology, 76,* 146-158.

Bunge, M. (1967). *Scientific research (Vols. 1 and 2).* Berlin: Springer.

Clore, G. L., Schwarz, N., & Conway, M. (1993). Affective causes and consequences of social information processing. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed.). Hillsdale, NJ: Erlbaum (in press).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Comrey, A. L. (1973). *A first course in factor analysis.* New York: Academic.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling.* London: Heinemann.

Dalkvist, J., & Rollenhagen, C. (1989). On the cognitive aspects of emotions: A review and a model. Report 703 from Department of Psychology, University of Stockholm.

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

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464-486.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. New York: Academic Press.

Frijda, N. H. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion, 1*, 115-143.

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

Garner, W., & Hartmann, D. P. (1984). On Markov dependence in the analysis of social interaction. *Behavioral Assessment, 6*, 229-236.

Gehm, T. L., & Scherer, K. R. (1988). Relating situation evaluation to emotion differentiation: Nonmetric analysis of cross-cultural questionnaire data. In K. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 61-77). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie [Measurement and modelling in psychology]*. München: Reinhardt.

Hays, W. L. (1973). *Statistics for the social sciences*. London: Holt.

Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.

Hubert, L. (1977). Kappa revisited. *Psychological Bulletin, 84*, 289-297.

Hunt, J. McV., Cole, M. W., & Reis, E. D. (1958). Situational cues distinguishing anger, fear, and sorrow. *American Journal of Psychology, 71*, 136-151.

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion, 3*, 81-123.

Kelly, G. (1955). *The psychology of interpersonal constructs*. New York: Norton.

Kemper, T. D. (1978). *A social interactional theory of emotions*. New York: Wiley.

Krech, D., & Crutchfield, R. S. (1958). *Elements of Psychology*. New York: Knopf.

Kukla, A. (1972). Foundations of an attributional theory of performance. *Psychological Review, 79*, 454-470.

Lazarus, R. S. (1968). Emotions and adaptation: Conceptual and empirical relations. In W. J. Arnold (Ed.), *Nebraska symposium on motivation* (Vol. 16, pp. 175-266). Lincoln: University of Nebraska Press.

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

Lazarus, R. S., & Smith, C. A. (1988). Knowledge and appraisal in the cognition-emotion relationship. *Cognition and Emotion, 2*, 281-300.

Lehrer, A. (1974). *Semantic fields and lexical structure*. Amsterdam: North-Holland.

Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.

Mandler, G. (1984). *Mind and body*. New York: Norton.

Manstead, A. S. R., & Tetlock, P. E. (1989). Cognitive appraisals and emotional experience: Further evidence. *Cognition and Emotion, 3*, 225-240.

Marx, W. (1982). Das Wortfeld der Gefühlsbegriffe. [The semantic field of emotion concepts]. *Zeitschrift für experimentelle und angewandte Psychologie, 29*, 137-146.

Mauro, R., Sato, K., & Tucker, J. (1992). The role of appraisal in human emotions: A cross-cultural study. *Journal of Personality and Social Psychology, 62*, 301-317.

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

Mees, U. (1985). Was meinen wir, wenn wir von Gefühlen reden? Zur psychologischen Textur von Emotionswörtern [What do we mean when we speak of feelings? On the psychological texture of emotion words]. *Sprache und Kognition, 1*, 2-20.

Meyer, W. U. (1973). *Leistungsmotiv und Ursachenerklärung von Erfolg und Mißerfolg [Achievement motive and causal explanation of success and failure]*. Stuttgart: Klett.

Neppl, R., & Boll, T. (1991). Analyse der Bedeutungsstrukturen alltagssprachlicher Emotionswörter. Gründzüge eines Verfahrens, exemplarische Anwendung, Implikationen für die Forschung zu spezifischen Emotionen [Analysis of the semantic structures of ordinary language emotion words]. *Sprache und Kognition, 10*, 85-96.

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models*. Homewood, IL: Irwin.

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

Pekrun, R. (1984). An expectancy-value model of test anxiety. In H. M. van der Ploeg, R. Schwarzer, & C. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 3, pp. 63-72). Lisse, The Netherlands/Hillsdale, NJ: Swets/Erlbaum.

Peters, R. S. (1970). The education of the emotions. In M. Arnold (Ed.), *Feelings and emotions* (pp. 187-203). New York: Academic Press.

Reisenzein, R., & Hofmann, T. (1990). An investigation of dimensions of cognitive appraisal in emotion using the repertory grid technique. *Motivation and Emotion, 14*, 1-26.

Reisenzein, R., & Hofmann, T. (1993). Discriminating emotions from appraisal-relevant situational information. Baseline data for structural models of cognitive appraisals. *Cognition and Emotion, 7*, 271-293.

Reisenzein, R., & Schönpflug, W. (1992). Stumpf's cognitive-evaluative theory of emotion. *American Psychologist, 47*, 34-45.

Roseman, I. J. (1979). *Cognitive aspects of emotions and emotional behavior*. Paper presented at the 87th Annual Convention of the APA, New York, Sept.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver, (Ed.), *Review of personality and social psychology* (Vol. 5, pp. 11-36). Beverly Hills: Sage.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition and Emotion, 5*, 161-200.

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

Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology, 60*, 37-47.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1, pp. 49-80). New York: Academic.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman, (Eds.), *Approaches to emotion* (pp. 293-317). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89-126). Dordrecht, the Netherlands: Kluwer.

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion, 7*, 325-355.

Schmidt-Atzert, L. (1981). *Emotionspsychologie [The psychology of emotion]*. Stuttgart: Kohlhammer.

Schmidt-Atzert, L., & Ströhm, W. (1983). Ein Beitrag zur Taxonomie der Emotionswörter [A contribution to the taxonomy of emotion words]. *Psychologische Beiträge, 25*, 126-141.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion Knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061-1086.

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology, 56*, 339-353.

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

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

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smolenaars, A. J., & Schutzelaars, A. J. H. (1986/1987). On "cognitive" semantics of emotion words: Solomon quasi-ecologically tested. *Journal of Semantics, 5*, 207-231.

Solomon, R. C. (1976). *The passions*. Garden City, NY: Anchor Press/Doubleday.

Stumpf, C. (1899). Über den Begriff der Gemüthsbewegung [On the concept of emotion]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 21*, 47-99.

Taylor, G. (1987). *Pride, shame, and guilt: Emotions of self-assessment*. Oxford: Oxford University Press.

Tesser, A. (1990). Smith and Ellsworth's appraisal model of emotion: A replication, extension, and test. *Personality and Social Psychology Bulletin, 16*, 210-223.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Weiner, B. (1982). The emotional consequences of causal attributions. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition* (pp. 185-209). Hillsdale, NJ: Erlbaum.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.