Will the Real Relationship between Facial Expression and Affective Experience Please Stand

Up: The Case of Exhilaration

Willibald Ruch

Department of Physiological Psychology

University of Düsseldorf

Running Head: FACIAL EXPRESSION AND AFFECTIVE EXPERIENCE

Abstract

It is hypothesized that the empirical correlation between facial expression and affective experience varies as a function of the correlational design used to compute the coefficients. Predictions about the rank order of five designs were derived based on two assumptions. Female subjects were placed into one of three alcohol conditions (no ethanol, low dose, high dose) and were exposed to 30 slides containing jokes or cartoons. The degree of rated funniness and overt behavior were intercorrelated using five different designs to analyze the same set of data. The results show that within-subject analyses yielded higher coefficients than between-subjects analyses. Aggregation of data increased the coefficients for within-subject analyses, but not for between-subject analyses. A cheerful mood was associated with hyperexpressiveness, i.e., the occurrence of smiling and laughter at relatively low levels of perceived funniness. It was demonstrated that low correlations between facial expression and affective experience may be based on several method artifacts.

<u>Will the Real Relationship between Facial Expression and Affective Experience Please Stand Up:</u>

<u>The Case of Exhilaration</u>

Emotions, like other concepts in psychology, are hypothetical constructs which are not directly observable, but are inferred from several indicators. The inference is typically based on behavior, physiological changes, or reports of subjective experience. Emotions vary in intensity and these intensity differences can be observed at all three levels. At least for the levels behavior and experience it is commonly assumed that intensity of the emotion is reflected equally well in these domains; <u>e.g.</u>, parameters of facial expression and affective experience are expected to show a high positive intercorrelation.

As regards the emotion induced by humor stimuli, however, empirical research has shown that while these manifestations of emotion are, indeed, positively related, the correlations are only low to moderate in strength. McGhee (1977) notes that the correlations between rated funniness and the amount of smiling or laughter typically range from .30 to .40, seldom exceeding the latter value. These low coefficients may have contributed to the failure of humor researchers to conceptualize the behavioral, physiological, and experiential responses to humor in terms of an emotion construct; rather, funniness ratings and smiling/laughter tend simply to be referred to as intellectual and affective measures of humor appreciation, respectively.

<u>The Emotion of Exhilaration</u>

Ruch (1990) recently proposed that exhilaration is the emotion elicited by humor. He argued that exhilaration satisfactorily accounts for the behavioral, physiological, and experiential changes typically occurring in response to such stimuli as tickling or laughing gas, as well as humor. In addition to describing the three levels of manifestation of exhilaration, and their relationship to one another, he has also described stimuli and situations which elicit exhilaration, as well as factors which facilitate or inhibit it. It was proposed that the term "exhilaration" be used according to its Latin root (hilaris = cheerful) to denote either the process of making cheerful or the temporary rise in cheerful state (Ruch, 1993). Thus, in contrast to common language, the "excitement" component is de-emphasized in the proposed usage of the term.

While exhilaration may be seen as a facet of the positive emotion of "happiness" (or "joy"), it is probably the facet most strongly aligned with laughter; whereas empirical studies of happiness/joy

rarely report its occurrence, laughter is an inevitable response category in humor studies. The concept of exhilaration fully incorporates, but also goes well beyond, what has traditionally been understood by the so-called "humor response"; i.e., the perception of a stimulus as being "funny." Basically, "humor response" is a concept of cognitive experience, regardless of whether it is or is not accompanied by smiling or laughter. Besides explicitly considering the levels of behavior and physiology, the concept of exhilaration claims with respect to the experiential level, that when people react to humor, their awareness is not restricted to perceiving the joke as funny; they are also aware of changes in feeling state, of physiological changes, and of their actions and action tendencies (Ruch, 1990). In partial support of this claim, it has been demonstrated that the degree of perceived funniness of jokes and cartoons is correlated with the degree of felt exhilaration (Ruch, 1981; Ruch & Rath, 1993). Indeed, the range of the coefficients (from .82 to .92 with a mean of .86) suggests, that these two judgements produce practically interchangeable results. However, the question of the extent to which measures of funniness or felt exhilaration correlate with exhilaration behavior remains to be answered.

Correlation between Smiling/Laughing and Funniness

McGhee (1977) drew attention to the problem of the low intercorrelation between funniness ratings and smiling/laughing and suggested that researchers "... include both measures as dependent variables and report the correlation obtained between them. This will provide a data base from which hypotheses may be drawn and tested regarding factors which influence the relationship between affective and intellectual measures of appreciation. It would also permit a cataloguing of the types of experimental manipulations to which each measure is sensitive" (p. 205).

In response to this suggestion, Ruch (1990) searched the humor literature for those articles which did report correlations between measures of smiling/laughter (using varying measurement approaches, such as observer judgments of degree of smiling or laughter, facial electromyography (EMG), and the Facial Action Coding System by Ekman & Friesen, 1978) and funniness ratings or other self-reports of positive affect. He found that across 25 different studies, the correlations ranged from about -.30 to a nearly perfect positive relationship. Three studies yielded negative correlations, and about half of the studies produced positive relationships below the .40 correlation

described by McGhee (1977). The important question, of course, is how can this wide range of correlations be explained?

Authors who obtained low coefficients or coefficients differing in size put forward hypotheses aimed at explaining these effects. For example, Branch, Fine, and Jones (1973) argued that the content of the humor stimulus makes an important contribution to the strength of this relationship. Among humor films which contained "tendentious" (i.e., sexual) content, they found a .63 correlation between funniness ratings and laughter; nontendentious films, however, yielded only a .30 correlation. Branch et al. (1973) also found that funniness ratings were more highly correlated with measures of smiling across a sample of subjects than measures of laughter were; low coefficients for laughter (ranging from -.31 to .41) were also found by Porterfield et al. (1988). Weaver, Masland, Kharazmi, and Zillmann (1985) argue that alcohol intoxication might be capable of impairing the correspondence between verbal reports and facial expressiveness, maybe because overt expressive behavior of the subjects, or the resulting feedback, is reduced or in some way altered by the effect of alcohol intoxication. This argument is based on the observation that the correlations between rated funniness of comedy films and several measures of facial expression were significant for the control group (.33 to .39), but were "substantially lower and non-significant" for low and high alcohol conditions. These and related hypotheses are a posteriori explanations and were never examined further.

The low coefficients also led to speculation about potential different meanings of the verbal reports and overt behavior. For example, it was proposed that smiling represents the intellectual, and smiling/laughing the affective appreciation of humor (Gavanski, 1986; Porterfield et al., 1988). Recently it was also claimed that amount of smiling reflects not the emotion, but the social context (Fridlund, 1991); this assertion was again based on the observation of a non-significant relationship between EMG-recordings of the contraction of the zygomatic major (i.e., the smiling muscle) and ratings of happiness.

Ruch (1990) argued that while situational and other social or subject factors may have some impact on the relationship between subjective judgments of funniness and the emotional reaction shown, methodological issues may hold the key to understanding the discrepancy generally found between these two aspects of humor appreciation. The size of the coefficients obtained may be

influenced by several factors, such as the nature of the assessment method (e.g., spontaneous judgments by an observer vs. use of a videotape for rating, use of psychophysiological measures), validity (e.g., control of method artifacts, separation of types of smiles) and reliability of the coding systems used, or the way the data are coded (e.g., simple judgments of presence or absence of smiling or laughter, discriminations of degree of smiling or laughter, or anatomically-based measures of facial activity involved in smiling/laughter). Also, the type of assessment of the affective experience might be crucial; rating scales may be more or less well anchored (e.g., defining the endpoints of the scale vs. defining each scale point clearly), the emotional quality might be more or less adequate (e.g., assessment of global happiness or of exhilaration, or amusement), and the frame for evaluating the stimuli can be stable (e.g., practice items preceding the actual stimuli allow for an emotional warm up and the establishment of a reference point for the funniness of the stimulus material to come) or changing. Further factors relate to the frequency (e.g., single vs. multiple measurements) and timing (e.g., immediate vs. delayed) of the assessment of affective experience and facial expressions.

While these considerations should influence the general level of the correlations obtained, Ruch (1990) argued that the design chosen to compute the correlations is also a major - if not the most important - determinant of the actual size of the coefficients. Thus, it was claimed that for a given set of data, the strength of the relationship between smiling/laughing and funniness judgements will vary in a predictable way depending on what data analysis strategy is applied. In other words, the experimenters' choice of a particular design predetermines the range of coefficients to be obtained, as well as the conclusions drawn about the strength of the relationship between judged funniness and smiling - or more generally - between facial expression and affective experience. Furthermore, it was argued that the inconsistent and low correlations obtained in prior studies are really an artifact based on violations of the requirements for the computation of correlations.

Five Ways to Compute the Correlations Between Smiling/Laughing and Judgements of Funniness.

Ruch's (1990) review of studies of the relationship between smiling/laughing and judged funniness showed that five different correlational designs - or data analysis strategies - have been used. In general, the computation has been based on two data matrices: one for verbal judgements and one for facial expression. In both matrices, n subjects give responses to m stimuli. The m

responses can be added (summed across stimuli) and form an individual's total appreciation score (row total). Conversely, the $\underline{n}$ responses to a stimulus can be added (summed across subjects) to form a total appreciation score for each stimulus (column total). The five designs differ with respect to whether correlations for data-pairs (e.g., smiling and judgements of funniness) are computed across columns, rows, or columns and rows simultaneously, and whether raw or aggregated data are used. Table 1 shows the five different designs.

_____

Insert Table 1 about here

_____

In Design A (between-subjects design for aggregated data), the correlation between the corresponding row totals (total appreciation scores for a given subject) of the two matrices is computed. One coefficient is obtained expressing the strength of relationship between subjects' perception of funniness of the total set of stimuli presented and the frequency of smiles (or laughter) displayed in response to them. A high coefficient indicates that subjects who smile more often also find the jokes more funny. In Design B (within-subjects design for aggregated data), the correlation between the corresponding column totals (total appreciation scores for a given stimulus) of the two matrices is computed. The resulting coefficient reflects the strength of relationship between the degree of funniness of the jokes (as perceived by all subjects) and the frequency of smiles induced by them. A high coefficient means that funnier jokes also elicit more smiles.

Designs C and D correspond to Designs A and B, respectively; however, they are based on raw rather than aggregated data. In Design C (between-subjects design for raw data), the correlations between the corresponding columns of the two matrices are computed separately for each joke, and are computed across all subjects. A high coefficient indicates that subjects who tend to find a particular joke funnier also tend to smile more in response to it. One coefficient is obtained for each stimulus, and the median of these $\underline{m}$ coefficients represents the strength of the relationship. In Design D (within-subjects design for raw data), the correlations between the corresponding rows of the two matrices are computed separately for each subject. They are computed across all stimuli, resulting in $\underline{n}$ coefficients (one for each subject). Again, the median can be used to represent the average strength of relationship. A high coefficient here shows that the more a particular subject tends to find the jokes funny, the more likely it is that he or she also smiles in response to them.

Finally, in Design E (across all stimuli and subjects design) the separation of subjects and stimuli is dissolved and the matrix is treated as a long (n times m) vector of data points. Thus, the correlation is computed across subjects (rows) and stimuli (columns) simultaneously, resulting in only one coefficient. A high coefficient means that in general (regardless of effects associated with subjects or stimuli), higher ratings of funniness are more likely than lower ratings to be accompanied by smiles.

All five designs have been used to analyze humor appreciation data, but a particular study has typically employed only one design; two were occasionally used, but no study has employed all five strategies. Researchers have not given reasons for choosing the particular design used; nor have they discussed the implications associated with them.

The Role of Design in Determining the Size of Coefficients: Two Premises and Four Deductions

What is the basis for assuming that the different designs have an impact on the size of the coefficients obtained? What is the typical rank order of the five designs with respect to the size of the coefficients they produce? Two basic assumptions permit the deduction of statements about the rank order of the designs.

First, differences between individuals affect the behavior-experience discrepancy more strongly than differences between stimuli. Thus, it is assumed that the variability in response patterns to different stimuli is lower than the variability in response patterns which exists among different people. Let's consider three hypothetical response patterns; one which represents the expected level of emotional expression for a given subjective experience, and two which deviate in the direction of hyper- and hypoexpressivity. All three response patterns are equally consistent, however. In the first case, affective experience and overt behavior are balanced; i.e., with increasing degree of felt emotion, the intensity of facial expression increases. In the case of humor, this means that there is no smiling when no funniness is perceived. With increasing funniness, however, a faint smile appears, then a strong smile, a laugh, and finally a strong laugh. Hyperexpressivity represents one kind of imbalance between behavior and experience; stronger forms of expression are shown at any given level of affective experience. For example, laughs occur at lower levels of funniness. In hypoexpressivity, there is a shift in the other direction; the intensity of the behavior is below what

might be expected from the level of affective experience. In the case of humor, this means that smiling or laughter only occur at very high levels of funniness.

What characteristics of the person can cause deviations from balanced expressivity? They include both temporary and habitual factors. For example, individuals in a cheerful mood smile even at jokes which are not very funny, while those in a more serious state produce the same intensity of smile only if the joke is very funny. On the other hand, some individuals habitually smile or laugh more easily than others do, although they do not differ in the intensity of felt affect. For example, one might expect that at comparable levels of judged funniness, extraverts laugh, ambiverts smile, and introverts do not even smile. However, such variations might also be a mere artifact resulting from the interindividually different use of rating scales.

What factors on the stimulus side may cause deviations from balanced expressivity? There are reasons to assume that different types of humor (e.g., sexual humor vs. puns) judged to have comparable levels of funniness may induce a disproportionately high or low level of overt responses. "Display rules," for example, can influence expressivity without influencing verbal judgments. In the presence of others, it might be appropriate to intensify, reduce, or even suppress smiling or laughter. There is no cause, however, to assume that different jokes will be accompanied by disproportionately stronger or weaker facial expressions if spontaneous expressions are not affected by display rules. While one humor category (e.g., sexual humor) may produce greater laughter than another (e.g., puns), this greater appreciation should be reflected in higher funniness ratings, as well as more laughter. However, it should not yield hyperexpressivity, as defined above. Thus, the first postulate claims that the degree of hypo- or hyperexpressivity which occurs in the case of reactions to humor is more likely the result of differences among people than differences among jokes or cartoons.

Deduction 1, then, is that Design D produces higher coefficients than Design C. For within-subjects correlations (Design D) the increasing degree of deviation from balanced expressivity might reduce the size of the coefficient due to floor or ceiling effects. For example, hyperexpressive subjects will not smile until reaching a certain level of affective experience. However, the same amount of imbalance (hyper- and hypoexpressivity) in a sample will reduce the coefficient more strongly if the correlation is computed across subjects (Design C). In the latter

case, for example, a judgement of a joke as moderately funny (e.g., a 4 on a 1 to 7 scale) could accompany no facial response, a smile, or a laugh in hypoexpressive, balanced, or hyperexpressive subjects, respectively.

The second assumption needs little explanation since it is known from measurement theory: aggregation of data may average out other sources of variation, and enhances reliability. The greater the number of items used to measure a construct, the more reliable the measurement gets. However, different effects can be expected for aggregation of data across stimuli and across subjects. When summing responses to the stimuli across subjects, the effects of differences among subjects average out and the differences between the stimuli get more reliable in both the funniness and the facial expression data. The degree of a person's balance or imbalance is present in all jokes, so it is kept constant. The differences among stimuli with respect to their exhilarating potential, however, are more reliably measured as an effect of the aggregation. Thus, it follows that Design B yields higher coefficients than Design D (Deduction 2); i.e., the aggregation of data for within-subjects designs increases the size of the coefficient (with the amount of increase being a function of the sample size).

Summing the subject's responses across the stimuli, however, has a different effect. On one hand, one can expect the aggregation of data to stabilize the variance due to different levels of appreciation of humor; i.e., the fact that some people tend to like humor more than others is more reliably assessed by many items than by one item only. However, repeated measurement also makes individual differences in the form of expressivity more reliable; the same degree of judged funniness of the cartoons or jokes might be associated with few smiles in hypoexpressives but with many facial responses in hyperexpressives. These differences should be more noticeable than in the case of responses to single stimuli, since - in individual jokes - both types of subject might fail to smile at low levels of funniness, but laugh at very high levels of funniness. Thus, one can assume that the aggregation of data in the across-subjects designs will not yield an increase in the size of the coefficients comparable to the one for the within-subjects designs. Whether there is an increase at all, no difference, or even a decrease depends on the relative size of the variance due to the form of expressivity and the variance due to general appreciation of humor. The greater the predominance of the former, the more likely the coefficient for A will be lower than the one for C;

conversely, the more the latter is predominant, the more likely the coefficient for A will be equal to, or even higher than, the one for C. Thus, no specific prediction about differences between Designs A and C can be made. However, it is clear that Design B will also exceed the coefficients of Design A (Deduction 3). Finally, since Design E incorporated both the effects of subjects and of stimuli, one can expect that the obtained coefficient will be between the ones for Design D and Design C (Deduction 4).

Given that the two postulates are correct, one can expect the deduced rank order of the designs when analyzing a particular data set with the amount of difference being affected by several factors, such as the number of data points aggregated, or the amount individual differences in expressivity present. It should be noted, however, that while the impact of the designs on the size of the coefficients obtained can and will be empirically demonstrated, this is not a case for the involvement of statistical inference. Different data analysis strategies cannot be statistically compared to each other, irrespective of how obviously they yield different results. The validity of the factors involved in the premisses (from which the differences among the design are deduced) can be examined, however, and the present study will examine one basis of the first postulate (assuming that the second postulate will be commonly subscribed to).

It was argued (Ruch, 1990) that the best way to demonstrate the impact of the designs on the size of the coefficients is to apply all data analysis strategies to the same set of data. While the 25 studies reviewed provided clear evidence for three of the predictions deduced, this review was limited by the fact that the studies compared differed not only with respect to the designs employed, but also with respect to other possible influential factors (e.g., assessment methods used, sample size, number of stimuli). Furthermore, the low number of studies employing Designs E and C, and the wide range of size of the coefficients (ranging from below zero to higher than .80) in Designs A, C, and D, did not allow for a reliable estimation of an average coefficient.

The present experiment.

The focus of the present article is the attempt to illuminate some of the mechanisms that gave rise to doubts that facial expression and affective experience are highly interrelated. Generally, the idea advanced in the present paper is that low correlations should not inevitably be interpreted as non-consistent response patterns or as evidence of a behavior-experience discrepancy. When low

coefficients have been obtained in the past, this might also have been due to the fact that an inappropriate design was employed in the data analysis or that method artifacts were not detected.

Thus, the <u>main</u> aim of the present study is to demonstrate using a single data set that the strength of the relationship between facial expression and affective experience varies in the predicted way as a function of the design used to compute the correlations. This phenomenon will be examined for all three experimental groups separately and for the total sample.

The <u>second</u> aim of the present study is to replicate the low coefficients reported for three variables and to clarify their causes. Specifically, it will be evaluated whether the effects of alcohol on the size of the coefficients reported by Weaver et al. (1985) is restricted to between-subjects designs. While it is expected that intake of alcohol significantly reduces the relationship between funniness and overt behavior in Design A (as found by Weaver et al., 1985), it is assumed that alcohol alters the balance between experience and behavior differently in different subjects <u>without</u> reducing the consistency of the response patterns. In other words, will be demonstrated that the hyper- and hypoexpressive subjects of the high alcohol group who are responsible for reducing the size of the coefficients in Design A, in fact have a highly consistent response pattern (i.e., yield high coefficients in Design D). Assuming that the view that alcohol primarily induces hyperexpressiveness is correct (see Weaver et al., 1985, who note that it is commonly assumed that alcohol makes one "laugh at poorer jokes") the number of hyperexpressive subjects will be higher in the alcohol groups than in the control group, whereas the number of hyperexpressives will not differ. Furthermore, it will be evaluated whether the coefficients are higher for tendentious forms of humor than for non-tendentious humor, as found by Branch et al. (1973). Finally, it will be evaluated why laughter results in lower correlations than smiling (Branch et al., 1973; Porterfield et al., 1988).

The <u>third</u> aim of the present study is to investigate one of the factors which formed the basis for the postulate that lower coefficients can be expected when the relationship between affective experience and facial expression is studied across subjects as compared to within subjects. It will be examined whether the degree of cheerful mood can account for the subjects' differences in expressivity; <u>i.e.,</u> whether subjects low and high in state cheerfulness are characterized by hypo- and hyperexpressiveness, respectively.

<u>Method</u>

<u>Subjects</u>

Sixty-three paid female non-psychology students volunteered to participate in this experiment. They were screened according to several criteria, such as medication, current health status, history of alcohol problems, possible pregnancy, and quality of vision. The videotapes of two subjects were partly unscoreable, leaving 61 subjects. Their age ranged from 19 to 30 years ($M =$ 22.9; SD = 3.0 years).

<u>Procedure</u>

<u>Overview</u>. A first meeting served to inform subjects about the aim and the course of the experiment. They were told that they were participating in a study examining the effects of alcohol intake on finger dexterity and that depending on which group they were in, they might receive a relatively low dose of alcohol. Their height and weight were determined and subjects were asked to avoid eating during the three hours prior to participating in the main experiment. Approximately one week later subjects returned for the main experiment lasting 90 minutes and were tested individually by a female experimenter. They first completed a mood inventory and a test measuring ability to concentrate. They then received under double blind conditions a beverage which they consumed within five minutes. Subsequently, they completed several dexterity tests and answered the mood scale again. After the period of induction of exhilaration they worked on the dexterity tests again.

Upon completion of the experiment the experimenter opened an envelope which contained the code for the beverages and informed the subject about the content of the beverage. Subjects in the experimental groups were given special attention while the effects of the alcohol wore off. None of the subjects reported any sign of discomfort. When debriefed, they were informed that the main aim of the experiment was to study facial responses to humor, and that the camera served this purpose rather than filming their dexterity. All subjects gave consent to have their tapes analyzed.

<u>Consumption of alcoholic beverage</u>. Subjects of all three groups consumed a beverage containing a mixture of peppermint water and grapefruit juice prepared by the experimenter of the first session. While the subjects of the control group (C) drank this mixture only, 0.22g and 0.39g

alcohol per kg body weight was added for the subjects of the low (E_1) and high (E_2) dose alcohol groups, respectively. For the latter two groups a 90% ethanol was used and was mixed with peppermint water on a 1:1 basis and filled up to 0.2l with grapefruit juice. These manipulations were expected to lead to blood alcohol levels of .4‰ and .7‰ (Forth, Henschler, & Rummel, 1975). For the sake of clarity, the two experimental groups are referred to as low and high alcohol groups in the present study, although these levels of blood alcohol should be regarded as low and medium doses, respectively; "high dose" is typically reserved for levels from 1.0‰ to 1.5‰ (Pandina, 1982). No high dose was used, since elevation of positive affect can mainly be expected for low and medium doses (Forth et al., 1975). The doses applied can be expected to be effective after 25 to 35 minutes (Forth et al., 1975; Ideström & Cadenius, 1968; McCollam, Burish, Maisto, & Sobell, 1980; Sidell & Pless, 1971). Hence the exhilaration-induction procedures began 30 minutes after consumption of the beverage. The application of the dexterity tests not only helped bridge this period of time; it should also have confirmed to subjects that the purpose of the experiment was the examination of dexterity under states of alcohol intoxication.

Assessment of cheerful mood. The degree of cheerful mood was assessed by means of the short version of the Eigenschaftswörterliste (EWL-K; Janke & Debus, 1978), a mood questionnaire aimed at providing a comprehensive assessment of mood state. The EWL-K contains a list of 123 mood-descriptive adjectives which are answered in a yes/no-format, and provides scores for 14 mood dimensions. There is no separate scale of cheerfulness; however, a measure of cheerful mood was derived by combining the six items of the "elation" scale which directly relate to cheerfulness. The EWL-K was administered twice to subjects; at the beginning of the experiment and immediately before the period of induction of exhilaration. It was expected that only degree of cheerful mood at the second administration would predict the degree of exhilaration induced.

Induction of exhilaration. Subjects sat in a comfortable chair in a slightly darkened part of the room and were presented one of two permutations of a set 35 black and white slides of jokes and cartoons at an interval of 25 seconds. The selection of humor stimuli was based on a taxonomy of jokes and cartoons (Ruch, 1992). The three categories of incongruity-resolution humor, nonsense humor, and sexual humor were represented by 10 slides each. This set was preceded by five "warming up" humor stimuli which were not used in any analyses.

Verbal responses. The subjects were instructed to rate all 35 stimuli according to six criteria. Two seven-point Likert scales assessed perceived funniness and aversiveness. The degree of funniness served as an indicator of the degree of exhilaration induced. Subjects also indicated in a yes/no-format whether they found the humor stimulus childish, tasteless, or boring, and whether they already had heard or seen the jokes or cartoons before.

Facial measurement. Measurements were made from color videotapes, which provided a close-up, head-on view of the subject's face. The camera was visible to subjects, but subjects in a pilot study, as well as in the main experiment, reported not having felt disturbed by the camera and that they believed that the camera was focused on their hand in order to assess dexterity. The measurements were based on the Facial Action Coding System (FACS; Ekman & Friesen, 1978). The FACS is an anatomically based, comprehensive, objective technique for measuring all observable facial movement. It distinguishes 44 action units (AUs). These are the minimal units that are anatomically separate and visually distinguishable. Five AUs were of relevance in the present study and their frequency, intensity (slight, moderate, extreme), symmetry/asymmetry and duration were coded according to the criteria given in the FACS manual. Three action units, AU12, AU6, and AU7 were expected to occur as part of both smiling and laughter. AU12 refers to the contraction of the zygomatic major, which pulls the lip corner up diagonally toward the cheekbone. AU6 and AU7 describe the actions of the outer and inner part of the orbicularis oculi muscle, respectively. The action of the former raises the cheek and causes crow's feet and wrinkles below the eye; the latter raises and tightens the lower eye lid. "Smiles" which contained AU13 or AU14 (alone or together with AU12) were not considered to represent positive affective responses to humor (Ruch, 1990). AU13 refers to the contraction of the caninus muscle which pulls the lip corner up sharply and puffs up cheeks. AU14 describes the dimplelike wrinkles at lip corners produced by the contraction of the buccinator muscle. Also, if AU12 was asymmetrical and shorter than 2/3 of a second, it was excluded from the analyses.

Additionally, the occurrence of laughter was coded. A laugh was defined as an initial forced exhalation, followed by a more or less sustained sequence of repeated expirations of high frequency and low amplitude, which may or may not be phonated as "ha-ha-ha." Episodes of a single audible

forced expiration occurring together with an AU12 formed the lower end of the intensity spectrum of events coded as laughter. However, a fully developed laughter pattern was generally observed.

The 1830 (30 stimuli times 61 subjects) responses were coded independently by two coders. In case of disagreement, they watched the relevant responses together until agreement was achieved. The general degree of agreement between the two coders based on 550 responses (composed of the five "warm up" jokes, as well as tapes of pilot subjects) was 93.1%.

Results

Impact of Design on the Size of the Coefficients

Product-moment correlations between judged funniness and frequency of AU12 were computed for each of the three groups separately, as well as for the total sample. All five correlational designs were used. The results are presented in Table 2.

_____

Insert Table 2 about here

_____

Table 2 shows that in all three experimental groups (as well as the total sample) the within-subjects designs (D and B) yielded higher coefficients than the between-subjects designs (C and A), both for raw and aggregated data[1] supporting Hypothesis 1 and 3, respectively. Furthermore, the aggregation of data in the within-subjects designs (B and D) yielded an increase of coefficients in all three groups (Hypothesis 2). The fact that the total sample yielded a higher coefficient than the three separate groups for Design B reflects the effect of the aggregation of data; the coefficient is based on the sum of data for 61, rather than 20 or 21 subjects. Accordingly, the difference between B and D was higher (.25) for the total sample than for the average (.17) of the three groups. Finally, in partial support of Hypothesis 4 the coefficient of Design E (across all stimuli and subjects) was between those for C and D, although closer to C in all cases except the high alcohol group.

The aggregation of data for the between-subjects designs (A and C) yielded inconsistent effects. While in the group of non-intoxicated subjects the coefficient for Design A (aggregated data) was slightly higher than the one for Design C (raw data), the coefficients decreased in the alcohol groups (and in the total sample).

Causes of Negative Coefficients

When computing correlations involving the designs for raw data, a minimum of three occurrences of AU12 was required for jokes (Design C) or subjects (Design D) to be included in the analysis presented in Table 1. This criterion excluded 11 subjects for Design D, and since three subjects had no response at all, the median coefficient is based on 47 subjects. The range of coefficients was from .39 to .89. In the case of Design C, each joke elicited more than three responses in the total sample. The coefficients for Design C ranged from .33 to .77. However, in the separate analyses of the three experimental groups, 6, 7, and 9 (out of 30) humor stimuli were excluded for the control group, low and high[2] alcohol groups, respectively.

Coefficients based on <u>fewer</u> responses were computed as well and showed tremendous variability, ranging from .14 to .85 for Design C, and from -.25 to .92 for Design D. While the low-frequency-based coefficients account for <u>all</u> of the low and negative coefficients, they do not necessarily have to be low; for example, the median of the excluded coefficients in Design D was .57. When only subjects with five or more responses were considered, no coefficient in Design D was below .50. Thus, it may well be that the negative or very low positive coefficients observed by Ruch's (1990) review of studies were derived from similarly skewed binary data.

<u>The Effect of Alcohol</u>

Table 1 shows that Weaver et al.'s (1985) findings can be replicated; they used Design A (between-subjects, aggregated data) and found that while rated funniness and facial expression are significantly correlated ($p < 0.01$) among non-intoxicated subjects, the coefficient for the high alcohol group (E_2) was nonsignificant. In the present sample, the difference between the control group and E_2 was very large (.37) and just failed to reach significance ($p = .065$, one-tailed)[3]. However, the fact that no such intoxication-based effects can be observed for the other four designs casts doubt on the proposition that alcohol intoxication really generally impairs the correspondence between verbal reports and facial expressiveness.

But why does Design A yield such a low correlation for the high alcohol group? Technically, low and high correlation scatter plots differ from each other by the disproportionate high representation of subjects off the main diagonal in the former. Given the present context, there must be subjects which either rate jokes high in funniness but smile infrequently or smile more often than expectable from the lower ratings. The critical question now is whether these subjects (who reduce the

coefficients for Design A in the high alcohol group) do indeed respond inconsistently to the humor stimuli or have a <u>consistent</u> response pattern (which is, however, altered in the direction of hypo- or hyperexpressivity).

Figure 1 presents the relationship between rated funniness and AU12-frequency for the control group and the high alcohol group. An expressiveness index was computed by dividing the frequency of AU12 by total funniness. This ratio was multiplied by six to correct for the differences in range of possible scores. Individuals more than one SD above and below the mean were considered to be hyper- and hypoexpressives, respectively.

_____

Insert Figure 1 about here

_____

Figure 1 shows that this index detects four hypoexpressive individuals in each of the two groups; thus, hypoexpressiveness is a factor lowering the coefficients produced by Design A irrespective of alcohol intake. Although the total response rate of these 8 subjects was too low as compared to degree of rated funniness, their individual response patterns were highly coordinated. Their few responses rather consistently were displayed in response to the jokes they judged funniest; hence all hyperexpressives yielded high coefficients in Design D (within-subjects, raw data).

There were three and six hyperexpressive subjects in the control and high alcohol groups, respectively. The 5 subjects with the lowest funniness ratings (below 30) in the high alcohol group do demonstrate the phenomenon of hyperexpressiveness nicely. Although they found the jokes only mildly funny (average of less than "1" on a 0 to 6 scale) they showed an average frequency of facial responses (<u>i.e.,</u> between three and six AU12 displays, $\underline{M} = 5.2$), and most of these responses were of higher intensity. Nevertheless, they (like the other hyperexpressives) have a very consistent response pattern. Their coefficients in Design D ranged from .71 to .92 ($\underline{M} = .81$) and are thus clearly above the average (see Table 1). Thus, the other type of unbalanced expressivity, namely hyperexpressiveness also does not allow one to draw the conclusion that alcohol impairs the correspondence between affective experience and facial expression. The difference is that while hyperexpressives begin to display smiles at funniness ratings of 0 to 2 (median = 1), the median of these funniness-"thresholds" for hypoexpressives is 5.

While without doubt alcohol does not lower the consistency of individual response patterns, the exact mechanisms causing the low coefficient of Design A in the high alcohol group remain unclear. Compared to the control group there were twice as many hyperexpressive subjects in E_2, but this difference is not statistically significant. Figure 1 demonstrates a formal cause; the variability of both AU12-frequency and rated funniness is reduced in E_2 as compared to C (although only the former is significant, $F[19,19] = 2.89$; $\underline{p} < .05$). Although hyperexpressiveness among subjects who do not judge the jokes to be very funny contributes to the reduction of variability of facial responses, it does not fully account for it.

<u>Does Laughter Really Fail to Reflect Humor Appreciation?</u>

The following pattern of correlations with funniness judgments was obtained for all experimental groups, and for all designs used: the coefficients for AU12 (lip corner puller) were always highest, followed by AU6 (cheek raiser), laughter, and AU7 (lid tightener). The respective coefficients for the total sample for Design E (across all stimuli and subjects) were .63, .52, .29, and .19.

Do different elements of exhilaration behavior indeed differ in their ability to reflect the amount of humor appreciation? Several considerations prohibit this conclusion. Firstly, laughter represents higher intensity levels of exhilaration than smiling does (Ruch, 1990); why should a more intense manifestation of exhilaration be a poorer indicator of the affective experience? Secondly, it is known (Ekman & Friesen, 1982; Ekman, Friesen, & O'Sullivan, 1988) that the presence of the action of the orbicularis oculi marks the felt happy smile; thus, higher coefficients should be expected for the joint action of AU12 and AU6 than for AU12 alone.

Thirdly and most importantly, it has not been considered yet that the maximal coefficient obtainable in point biserial correlations depends on the distribution of the binary variable; <u>i.e.</u>, the relative frequency of the AUs. The form of the relationship between relative frequency and maximal coefficients can be described as an inverted-U function; the more the relative frequency deviates (in both directions) from .50, the lower the maximal possible size of the coefficient. Whereas AU12 occurred relatively often (432 cases), the other behaviors were displayed less frequently, the order of occurrence being the same with which the coefficients decrease (AU6 = 258, laughter = 71, and AU7 = 31). In test theory an index has been advanced to allow an

evaluation of item-total correlations for items with different endorsement frequency ("Selektionskennwert"; Lienert, 1969). If this index (division of the coefficient by twice the variance of the binary variable) is computed, the resulting values of AU12 (.75), AU6 (.74), laughter (.76), and AU7 (.73) do not differ from each other. Thus, it is safe to assume that smiling and laughter equally well signify exhilaration.

Effects of Type of Humor?

The use of three humor categories in the present experiment allows for a test of the hypothesis advanced by Branch et al. (1973) that the correlation between judged funniness and overt behavior will be higher for tendentious than for non-tendentious humor. Separate analyses were performed for the humor categories of incongruity-resolution, nonsense, and sexual humor. No systematic effect could be observed. First, the differences in size of correlation were not consistent across the designs; i.e., depending on the design, incongruity-resolution humor (Designs B and E), nonsense (Design C), or sexual humor (Design A) were highest. Second, the differences in size of coefficients were generally negligibly small. Third, where differences between the three humor types existed, they were probably due to the different response frequencies; i.e., the order among the humor categories with respect to the size of the coefficients and the size of the means was identical.

Does Cheerfulness Affect Expressiveness?

In order to determine whether the degree of cheerfulness of one's mood can account for individual differences in expressivity, two types of analyses were undertaken. First, the correlations between subjects' degree of cheerfulness (i.e., the score in the subscale of the EWL-K, at the second administration) and funniness "threshold" for facial expression (i.e., the minimal funniness rating which was accompanied by an overt facial response) were computed. The coefficients were computed across the whole sample, since a Kruskal-Wallis analysis showed that the three groups did not differ from each other with respect to degree of cheerfulness ($p = .76$) or threshold of AU12 ($p = .78$). The rank order correlation obtained between degree of cheerful mood and the funniness threshold for AU12 ($R (58) = -.27$, $p < .05$) confirms that subjects in a more cheerful mood show facial responses at lower minimal levels of rated funniness than do less cheerful subjects.

Second, the ability of cheerfulness to predict degree of judged funniness and facial behavior was compared. Whereas the coefficient for funniness (.21, ns) was not significant, cheerful mood correlated significantly with the frequency (.33, p <.01) and intensity (.29, p < .05) of AU12, the maximal response[4] shown (.41, p < .01), the joint action of AU12 and AU6 (.28, p < .05), and the frequency of laughter (.40, p < .01). Thus, subjects who were high and low in cheerfulness differed with respect to the frequency and intensity of facial behavior shown, but not in terms of their judgments of funniness. Identical conclusion can be drawn from the analysis of non-intoxicated subjects; the coefficients range from .33 for judged funniness (p = .15) to .59 for laughter (p < .001) and .61 for the maximal response shown (p < .01).

Thus, states of high cheerful mood go along with hyperexpressivity, and low cheerful mood goes along with hypoexpressivity. These effects were only found for the assessment of cheerful mood immediately before the beginning of the induction of exhilaration. The degree of cheerfulness at the beginning of the experiment did not significantly predict verbal or facial responses to humor. The average correlation between the two testing times was .67 (ranging from .29 in E_2 to .90 in the control group). Also, although the cheerfulness measure was composed of a subset of the items of the elation scale, the coefficients for the total elation scale were consistently lower than those obtained for the cheerfulness items. This justifies the use of a separate cheerfulness scale, and signifies the importance of this concept in research on exhilaration.

Discussion

The aim of the present study was to draw attention to the fact that the size of the correlation coefficient obtained between measures of facial expression and affective experience is influenced by several methodological factors. Low coefficients might be based on artifacts. The failure to consider them has led to unjustified conclusions about the nature and strength of the relationship between funniness and smiling, and perhaps even about facial expression and affective experience in general. An experiment was conducted which attempted a) to replicate the empirical basis on which several such conclusions are based, and b) to demonstrate that these conclusions are not justified.

It was argued that the approach used to compute the correlations constitutes the major neglected factor; even for a given data set it produces systematic variations in the size of the coefficients to be obtained. Starting from two premises, four statements about the relative size of the coefficients produced by the different data analysis strategies were deduced. All three experimental groups as well as the total sample were used to illustrate these predictions. Combining the four deductions (D > C; B > A; B > D; D > E > C) one arrives at B > D > E > C being the typical rank order of the data analysis strategies regarding the size of the coefficients to be obtained for them. Design A cannot be unambiguously located in this sequence. It is certainly below B and most likely below D, but whether it falls below or above C depends on whether - and to what extent - factors which increase the amount of variation on the hypo- vs. hyperexpressiveness dimension are the present in that data set .

Differential alcohol effects seem to be among such sorts of variance. While for the group of non-intoxicated subjects the difference between the highest and lowest coefficient was .25, this range increased to .59 in the high alcohol group! This was almost exclusively due to the reduction of the coefficients obtained for strategy A. As in the study by Weaver et al. (1985), the coefficient was not significantly different from zero in the high alcohol condition. Additionally, the present study verified a significant decrease as compared to the control group. However, the emergence of this phenomenon seems to be contingent on two factors: analysis *across* subjects and data *aggregation*. Only the across subjects analysis is affected by individual differences in type of expressiveness (between subjects analyses keep them constant) and data aggregation is needed to make this variance more reliable. The importance of the latter can be estimated from the fact that no reduction was observable for strategy C, where the coefficients were also computed across subjects. The nature of this differential[5] alcohol effect could not be substantiated, however. The doubling of the number of hyperexpressives was not significant. It appears that for a verification of this hypothesis a larger sample size is needed, or the same subjects would have to be studied in both intoxicated and non-intoxicated states using parallel sets of humor. The latter strategy would also allow one to study whether alcohol leads to an increase in overt behavior, or to a decrease in perceived funniness, or both. Whatever the nature of this effect turns out to be, the present study

clearly contradicts the postulate that alcohol consumption lowers the *coherence* between facial expression and affective experience.

Thus, the same set of data might suggest quite opposite conclusions, depending on which correlational design is used to analyze the data. While the span of coefficients in the present study can be regarded as extraordinarily high (ranging from .28 to .96!), a still considerable range can be expected for control group conditions, as was verified in two replication samples (Ruch, in preparation). In the replication study most similar to the present one (80 females, 42 jokes and cartoons, EMG-recordings of three facial muscles) the correlations between judged funniness and activity of the zygomatic major were .54, .89, .56, .63, and .58 for Designs A through E, respectively. Thus, the range of coefficients was .35.

Neglecting the impact of the data analysis strategy is not the only way of arriving at wrong conclusions about the relationship between verbal evaluations and facial reactions to humor on the basis of method artifacts. The present study quite clearly demonstrated another pitfall; namely, the comparison of coefficients for variables of different skewness or response frequency. The analyses showed that is not justified to conclude that the different elements of exhilaration behavior are correlated with judged funniness to different extents, despite the fact that the uncorrected coefficients suggested such an interpretation[6]. Similarly, the present study shows that it is safe to assume that the type of humor does *not* affect the size of the coefficients obtained. One has to keep in mind, however, that the real strength of the relationship for less frequently occurring responses is not underestimated only in the case of non-aggregated binary data (i.e., strategies C, D, and E) but also for aggregated data. Even after data aggregation, less frequently occurring responses remain less reliable than the more frequent responses, or, in the case of an insufficient number of data points, are more skewed than the latter. For example, the declining size of coefficients for AU12, AU6, laughter, and AU7 could also be observed for strategies A (.55, .42, .20, and .03) and B (.96, .89, .71, and .67).

The results obtained from applying the correction formula were supported by the previously mentioned replication study (Ruch, in preparation) in which the actions of the zygomatic major and the orbicularis oculi were assessed with the help of facial electromyography. The frequency of actions of the two muscles was about the same, so the coefficients could not have been biased by

this factor. Indeed, without any correction, the coefficients obtained for the zygomatic major and orbicularis oculi did not differ substantially in any of the five designs.

However, it is not only *methodological* factors which lower the relationship between judged funniness and smiling/laughing. The present study investigated one of the factors forming the basis for the postulate that lower coefficients can be expected when the relationship between affective experience and facial expression is studied across subjects as compared to within subjects. It was confirmed that degree of cheerful mood is associated with individual differences in expressivity. A cheerful state indeed seems to predispose to hyperexpressiveness; i.e., more cheerful individuals displayed smiling and laughter at lower levels of perceived funniness than less cheerful persons did. Thus, heterogeneity in mood level is a factor selectively affecting the across subjects analysis. It can be expected that an experimental control of mood variation would elevate the coefficients for strategies A and C.

What is, then, the 'real' relationship between affective experience and facial expression of exhilaration? Obviously it is much higher than it has previously been considered to be. First of all, even the lowest coefficient found for the non-intoxicated group (.61) is far beyond the .40 which was considered to be the upper limit of the range of coefficients typically obtained in humor studies (McGhee, 1977). This was mainly due to the fact that the optimal conditions for *generally* enhanced coefficients mentioned above were chosen. Furthermore, the present experiment identified several (method- and content-related) factors which *selectively* cause low coefficients even when the actual coherence of affective experience and facial expression is quite high. It is evident that the different data analysis strategies are likely to underestimate the strength of the relationship in *different* ways. The coefficients yielded by the between-subjects designs A and C are lowered by the presence of variance subsumed under the postulated dimension of hyper- and hypoexpressiveness. The greater the variation of subjects' response pattern on this dimension, the greater the differences between within-subjects and between-subjects designs. Additionally, Design A seems to be more susceptible to such influences than Design C, since aggregation of data enhances the reliability of individual differences in expressivity. While Ruch's (1990) review of studies showed that Design A is the one most frequently used, the present findings suggest that its sole use is not recommendable since it suggested an alcohol-induced behavior-experience

discrepancy where in fact there was no such discrepancy. The use of strategy A in combination with D (within-subject, raw data), however, would provide a sensitive indicator for the presence of such phenomena. While the within-subjects designs do take care of a lot of problems, such as differential expressivity, or the fact that some measurements are not comparable across subjects (e.g., absolute EMG-amplitudes), they do challenge the assumption of independence of measurements and thus prohibit the use of tests of significance (which are not necessarily needed, however). Coefficients of design D may be lowered for extreme hyper- and hypoexpressive subjects due to the deviation from linearity of the regression[7] (for continuous data) or the reduced maximal size of the coefficients (for binary data). These and others effects appear to be averaged out for Design B (within-subjects, aggregated data), especially for a higher number of subjects. The coefficients obtained for strategy E (across all stimuli and subjects) are lowered by the same factors - albeit to a lower extent - that affect designs C and D. Finally, irrespective of the chosen data analysis strategy, lowered coefficients will be obtained for low frequency behavior and particularly so in the analysis of binary raw data (i.e., strategies C, D, and E).

Since researchers have not yet given attention to the factors described in the present article, the conclusion commonly drawn that judgments of funniness show only a low positive relationship to overt expressions of affect is clearly premature, and probably without any foundation at all. On the contrary, it appears to be safer to assume that responses to humor can be described by a single emotion construct; both smiling/laughter and the judged funniness can be considered to indicate the same affective changes in the organism.

It appears recommendable to consider these factors also when studying the relationship between affective experience and facial expression in other emotions. Likewise, it might be of interest to consider the role of correlational designs in other areas of psychology where such dissociation-effects are discussed, for example, in research on psychopathology, or effects of psychoactive drugs, or in the study of attitude-behavior relationships.

References

Branch, A. Y., Fine, G. A., & Jones, J. M. (1973). Laughter, smiling and rating scales: An analysis of responses to tape recorded humor. Proceedings of the 81st Annual Convention of the American Psychological Association, 8, 189-190.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system (FACS). Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. Journal of Nonverbal Behavior, 6, 238-252.

Ekman, P., Friesen, W. V. & Ancoli, S. (1980). Facial signs of emotional experience. Journal of Personality and Social Psychology, 39, 1125-1134.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. Journal of Personality and Social Psychology, 54, 414-420.

Forth, W., Henschler, D., & Rummel, W. (1975). Allgemeine und spezielle Pharmakologie und Toxikologie [General and special pharmacology and toxicology]. Mannheim: Bibliographisches Institut.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. Journal of Personality and Social Psychology, 60, 229-240.

Gavanski, I. (1986). Differential sensitivity of humor ratings and mirth responses to cognitive and affective components of the humor response. Journal of Personality and Social Psychology, 51, 209-214.

Godkewitsch, M. (1976). Physiological and verbal indices of arousal in rated humour. In A. J. Chapman, & H. C. Foot (Eds.), Humour and laughter: Theory, research and applications (pp. 117-138). London: Wiley.

Ideström, C. D., & Cadenius, B. (1968). Time relations of the effects of alcohol compared to placebo. Psychopharmacologia, 13, 189-200.

Janke, W., & Debus, G. (1978). Die Eigenschaftswörterliste (EWL) [The adjective word list]. Göttingen: Hogrefe.

Lienert, G. A. (1969). Testaufbau und Testanalyse [Test construction]. Berlin: Beltz.

McCollam, J. B., Burish, T. G., Maisto, S. A., & Sobell, M. B. (1980). Alcohol's effects on
physiological arousal and self-reported affect and sensations. Journal of Abnormal Psychology, 89,
224-233.

McGhee, P. E. (1977). Children's humour: A review of current research trends. In A. J. Chapman, &
H. C. Foot (Eds.), It's a funny thing, humour (pp. 199-209). Oxford: Pergamon Press.

Pandina, R. J. (1982). Effects of alcohol on psychological processes. In E. Gomberg, H. White, & J.
Carpenter (Eds.), Alcohol, science, and society revisited. Ann Arbor: University of Michigan
Press.

Porterfield, A. L., Mayer, F. S., Dougherty, K. G., Kredich, K. E., Kronberg, M. M., Marsee, K. M.,
& Okazaki, Y. (1988). Private self-consciousness, canned laughter, and responses to humorous
stimuli. Journal of Research in Personality, 22, 409-423.

Ruch, W. (1981). Witzbeurteilung und Persönlichkeit: Eine trimodale Analyse. [Humor and
personality: A three-modal analysis]. Zeitschrift für Differentielle und Diagnostische Psychologie,
2, 253-273.

Ruch, W. (1990). Die Emotion Erheiterung: Ausdrucksformen und Bedingungen [The emotion of
exhilaration: Forms of expression and eliciting conditions]. Unpublished habilitations thesis,
Department of Psychology, University of Düsseldorf, Germany.

Ruch, W. (1992). Assessment of appreciation of humor: Studies with the 3 WD humor test. In C. D.
Spielberger, & J. N. Butcher (Eds.), Advances in personality assessment (Vol. 9, pp. 34-87).
Hillsdale, NJ: Lawrence Erlbaum Associates.

Ruch, W. (1993). Exhilaration and humor. In: M. Lewis, & J. M. Haviland (Eds.), The handbook of
emotion (Chapter 42, pp. 605-616). New York, NY: Guilford Publications.

Ruch, W., & Rath, S. (1993). The nature of humor appreciation: Toward an integration of perception
of stimulus properties and affective experience. Humor: International Journal of Humor Research.

Sidell, F. R., & Pless, J. E. (1971). Ethyl alcohol: Bloodlevels and performance decrements after oral
administration to men. Psychopharmacologia, 19, 246-261.

Weaver, J. B, Masland, J. L., Kharazmi, S., & Zillmann, D. (1985). Effect of alcoholic intoxication
on the appreciation of different types of humor. Journal of Personality and Social Psychology, 49,
781-787.

Author Notes

Footnotes

[1] It should be noted that this result can not be explained by a greater degree of variability between stimuli than between subjects. On the subjects's side, there were 3 with no single smile and 5 subjects smiled in more than half of the 30 times. On the variables side, all jokes had at least three facial responses and there was only one stimulus eliciting a response in almost half (30 out of 61) of the subjects.

[2] The elimination of this larger number of subjects the high alcohol group is responsible for the discrepant result that the coefficient for C exceeds the one for E.

[3] This difference was significant, however, for the intensity of AU12 (C: r = .71, p < .001; E_2: r = -.01 , ns; diff = .72, p < .01) and the frequency of the combined action of AU12 and AU6 (C: r = .69, p < .001; E_2: r = .12 , ns; diff = .57, p < .05).

[4] The maximal response was represented by the highest level of facial enjoyment coded for a subject. This five-point index covered "no response" (= 0), three intensity levels of AU12-smiles, and laughter (= 4).

[5] A *general* alcohol effect (e.g., if all subjects became equally hyperexpressive) would not reduce the coefficients of strategy A.

[6] The results found by Branch et al. (1973) might have been affected by the same factors; smiling was three times more frequent than laughter.

[7] Many thanks to the reviewer pointing this out.

Table 1

Designs used to Correlate Affective Experience and Overt Behavior

DESIGN	Correlated across	Data used	Aggregated across	Number of resulting coefficients
A:	n Subjects	Total scores	Stimuli	1
B:	m Stimuli	Total scores	Subject	1
C:	n Subjects	Raw scores	n.a.	m
D:	m Stimuli	Raw scores	n.a.	n
E:	n Subjects and m Stimuli	Raw scores	n.a.	1

Note. A = between-subjects design for aggregated data. B = within-subjects design for

aggregated data. C = between-subjects design for raw data. D = within-subjects design

for raw data. E = across all stimuli and subjects design. n.a. = not applicable

Table 2

Correlations between judged funniness and appearance of AU12

	Design				
	A	B	C	D	E
Total sample	.55	.96	.61	.71	.63
No alcohol	.65	.86	.61	.74	.62
Low alcohol	.54	.92	.63	.69	.64
High alcohol	.28	.87	.62	.72	.60

Note. A = between-subjects design for aggregated data. B = within-subjects design for aggregated data. C = between-subjects design for raw data. D = within-subjects design for raw data. E = across all stimuli and subjects design. The results for Designs C and D are median values of several correlations.

Figure Captions

<u>Figure 1</u> Scatterplots of the total scores for judged funniness and AU12-frequency for the control

group (a) and the high alcohol group (b). (▲ : hyper-, O: normo-, and ▼: hypoexpressive subjects)