COGNITION AND EMOTION, 1994, *8* (3), 201–229

Coherence between Expressive and Experiential Systems in Emotion

Erika L. Rosenberg and Paul Ekman

Human Interaction Laboratory, University of California, San Francisco, USA

In order to assess the extent of coherence in emotional response systems, we examined the relationship between facial expression and self-report of emotion at multiple points in time during an affective episode. We showed subjects brief films that were selected for their ability to elicit disgust and fear, and we asked them to report on their emotions using a new reporting procedure. This procedure, called *cued-review*, allows subjects to rate the degree to which they experienced each of several categories of emotion for many locations over the time interval of a stimulus period. When facial expressions and reports of emotion were analysed for specific moments in film time, there was a high degree of temporal linkage and categorical agreement between facial expression and self-report, as predicted. Coherence was even stronger for more intense emotional events. This is the first evidence of linkage between facial expression and self-report of emotion on a momentary basis.

INTRODUCTION

We endorse the view that emotions involve *patterns of responses* that have evolved for their ability to organise disparate bodily systems to respond efficiently to critical environmental events. This position—one version of evolutionary emotion theory—postulates different patterns of responses for each emotion, and implies a coherence among behavioural, physio-

Requests for reprints should be sent to Erika L. Rosenberg, Human Interaction Laboratory, University of California, San Francisco, CA 94143 USA (e-mail: erikar@itsa.ucsf.edu).

This research was supported in part by a National Science Foundation Graduate Research Fellowship to Erika L. Rosenberg and a Research Scientist Award from the National Institute of Mental Health (MH 06091) to Paul Ekman. We are grateful to Gregory Feist, Barbara Fredrickson, and Lisa Rosenberg for their insightful comments on earlier versions of this manuscript. We also thank Mark Frank for reliability coding.

logical, and subjective systems (Darwin, 1872/1965; Ekman, 1977, 1992; Levenson, 1988; Plutchik, 1962; Tomkins, 1962).

Tomkins (1962, pp. 243–244) wrote that affects are sets of organised responses that "are capable when activated of simultaneously capturing such widely distributed organs as the face, the heart, and the endocrines and imposing on them a specific pattern of correlated responses". In the empirical literature on emotion, evidence is just beginning to appear that there is cohesion among different emotion response domains. There has been some evidence of coherence between facial expression and self-report of emotion (Ekman, Friesen, & Ancoli, 1980) and for coherence between expression and physiology (Davidson, Ekman, Saron, Senulis, & Friesen, 1990; Ekman, Davidson, & Friesen, 1990; Levenson, Ekman, & Friesen, 1990). Such research, however, has not looked at correspondences among emotional response systems at specific points in time, but rather has evaluated correspondences among response measures from aggregated occurrences of particular emotions across blocks of time. If Tomkins was right, congruence between systems should be evident also on a more momentary basis. During an emotion, changes in expression, physiology, and experience should correspond temporally and categorically. Facial signs of anger should occur with the subjective experience of anger and anger-specific physiological patterns.

In contrast to previous research that has looked at overall correspondences in emotional response systems, we sought to determine whether facial expression and subjective measures of emotion occur together on a momentary basis during specific emotional episodes. Such congruence between measures is a minimum requirement for the type of coherence presupposed by our view of emotion. By *coherence* we mean integrated, concurrent, emotion-specific changes in different emotion response domains.

SOURCES OF COHERENCE

Our aim was not to demonstrate *why* coherence should occur between facial and subjective response systems but rather to determine the extent to which coherence does occur. It is important to note, however, that the mechanisms that are potentially responsible for congruence in emotional response systems are a matter of debate. Our view is that coherence is due to the central organisation of emotion in the brain (after Darwin, 1872/1965; Tomkins, 1962). This might be called *internally driven* coherence. Alternatively, observable coherence in emotional responding is consistent with facial-feedback theory (e.g. Laird, 1974). If peripheral stimulation can instantaneously feedback to the brain to cause emotional experience, then facial expression and reports of experience would appear to be congruent. This type of coherence might be called *externally driven*. There

is research to partially support both views, but none to distinguish among various explanations for coherence between expression and self-report (cf. Ekman et al., 1990; Davidson et al., 1990 for evidence of internally driven coherence, and Kleck et al., 1976; Lanzetta, Cartwright-Smith, & Kleck, 1976 for evidence for facial-feedback theory).

Another view is that there are inverse relationships among response systems in emotion. Jones' (1950) internalisation-externalisation hypothesis, which posits that affect inhibited overtly will be manifested physiologically and vice versa, is a well-known example of such a perspective. Such negative correspondences have been observed in normal human adults (e.g. Lanzetta & Kleck, 1970). Buck (1980) demonstrated, however, that the observed direction of the relationship between facial and physiological response measures of emotion may depend on analytical approach. Between-subjects analyses tend to reveal negative correlations between behaviour and physiology, whereas within-subjects correlations tend to be positive (although low to moderate in strength). This distinction has recently been further substantiated and developed (Cacioppo et al., 1992). Negative correspondences have *not* been demonstrated at the level of the individual in normal subjects, however.

The debate over the sources or extent of emotional response coherence has not been resolved by studies of the phenomenon, largely because the issue has not been addressed with sufficient methodological rigour. In the next few pages, we outline the major methodological issues that complicate the study of coherence and offer a new technique that addresses some of these problems.

Methodological Problems in the Study of Coherence

Evaluating Coherence at Multiple Points in Time. Although facial behaviour or facial muscle activity has been obtained continuously in previous studies, researchers have summarised or aggregated facial behaviour—whether measured in terms of observable expressions or EMG-recorded muscular activity—over the time period of each emotion episode in order to render it statistically comparable to single, retrospective emotion reports (Brown & Schwartz, 1980; Ekman et al., 1980, 1990). The correlations between summarised facial expression and single self-reports of emotion have ranged from 0.35 to 0.55, indicating congruence between these systems (Ekman et al., 1980). Correspondence between aggregate measures of expression and report does not tell us whether the two response systems cohere at specific points in time, however.

According to Levenson (1988), aggregation across long periods of time may compromise dynamic information about the emotion response.

Averaging across a period that may contain (p. 30) "a number of different emotions, as well as periods of no emotion" can obscure "patterning associated with a single emotion". On the other hand, some might argue that the relationship between affective facial behaviour and subjective experience would disappear at the momentary level, if congruence is an emergent property of aggregated episodes. It may be that for any given emotion there is no correspondence between facial and subjective responses, but rather after several distinct emotional events a pattern of congruence appears. Personality research on the stability of behaviour (Epstein, 1980) certainly indicates that intra-individual consistencies between behaviours can best be appreciated on a cumulative basis.

Special Problems of Measuring Subjective Experience

Measures of emotional experience have to be *elicited*, whereas measures of facial behaviour and physiology are *emitted* either continuously or sporadically throughout an emotional episode. This fundamental difference in the experimenter's access to the various emotional response systems entails two potential problems in the measurement of emotional experience: the problem of symbolic representation and the problem of continuity.

1. *The Problem of Symbolic Representation.* Measures of the subjective experience of emotion are fundamentally different from behavioural and physiological measures, in that they must be filtered through consciousness. Some processes may be truly inaccessible to conscious report, however (Nisbett & Wilson, 1977). Although signs of emotion may be manifest behaviourally or physiologically, people may not always be aware that they are undergoing emotion, or they might be aware of feeling an emotion but not be certain of which emotion it is. Even if people are clearly aware of the emotions they are feeling, they may choose to censor them (cf. Ekman, 1985).

Once people know they are feeling an emotion and choose to reveal their experience, that experience must by symbolically represented. This further reduces the comparability among response systems. Self-ratings of emotional experience may carry with them the baggage of error variance related to self-reports in general, such as acquiescence (Campbell, Siegman, & Rees, 1967), social desirability (Crowne & Marlowe, 1964), and demand characteristics (Orne, 1962).

There are a number of ways to have people symbolically represent their subjective experiences of emotion. Subjects could be asked to choose

colours or textures that best represent their emotions (Block, 1957) or to use rating dials or joysticks to quantitatively rate the degree of their positive and negative feelings (Levenson & Gottman, 1983, 1985). Most often, however, experimenters request a verbal report. Although offering more precision, this method imposes language on an experience that may not occur in verbal form. This problem is further complicated by the fact that most researchers do not allow subjects to use words of their own choice; instead they prescribe a specific set of words and/or a set of numbers that subjects must use to describe their emotional experiences (e.g. Ekman et al., 1980, 1990).

Another potential problem in the measurement of emotional experience is memory reconstruction. Usually, in studies of the relationship between facial expressions and/or physiological measures with subjective measures of emotion, single self-reports have been obtained after an emotional episode (Brown & Schwartz, 1980; Ekman et al., 1980, 1990). Several distinct emotions can occur even during a brief stimulus period, but we do not know which aspects of these emotional experiences are captured in single, retrospective reports. They may overemphasise the most recent or the most intense elements in the experience, or they may result from an averaging across the varied experience within a period. Retrospective emotion reports may involve the formation of "global impressions", which are vulnerable to potential influences of selection, recollection, and aggregation (cf. Reis & Wheeler, 1991). Studies of daily emotional experience suggest that relative to momentary reports, retrospective accounts of daily emotion overemphasise the intensities of emotions and show preference for negative over positive emotions (Thomas & Diener, 1990). Similar biases have been observed in research in mood and menstrual symptom reporting (Englander-Golden, Chang, Whitmore, & Dientsbier, 1980). Thus, comparisons of single reports of reconstructed experience with behaviour and/or physiology obtained during a spontaneous experience may yield less than perfect correspondences.

2. *The Problem of Continuity.* It is rare that a stimulus event calls forth one emotion. More commonly, multiple, brief emotions unfold over time, even over periods lasting for only a minute or two (Ekman, 1984). Therefore, to establish evidence for coherence between emotional response systems, one must examine congruence among measures of specific emotional events at particular points in time. Facial expression and physiological changes can be monitored continuously during an emotional episode. Concealed cameras can be used to videotape spontaneous facial expression of emotion, which can then be systematically scored (using a coding scheme such as the Facial Action Coding System; Ekman & Friesen, 1978) or facial muscle activity can be continuously monitored

using electromyography (EMG). Changes in several types of autonomic or central nervous system activity can be recorded online, throughout an emotional episode. There is no way to constantly monitor subjective experience of emotion that is analogous to the methods available for these other response systems.

The methods that do exist for measuring subjective experience on a relatively continuous basis may introduce artefacts. Thought sampling, wherein subjects report on their experiences while they are undergoing an emotional event, is one option. The spontaneous, online nature of this procedure is advantageous, but the constant demand to report on experience may to an unknown extent interfere with the degree to which the subject becomes emotionally responsive to the stimulus. Additionally, speaking aloud one's feelings may increase self-consciousness and the sense of being observed, which may alter affect as well as expressivity (cf. Kleck et al., 1976).

Some investigators have had subjects provide continuous ratings of their affect during a video replay of their emotional interactions (Kagan, Krathwohl, & Farquhar, 1965; Levenson and Gottman, 1983, 1985). In Levenson and Gottman's (1983, 1985) video-recall technique, subjects use a dial to rate continuously their recalled affect. Turns in one direction convey the degree to which the interaction evoked positive emotion; turns in the other direction reflect negative emotion. Subjects make these ratings after the emotional event while watching a video of their emotional interaction, although recently investigators have had some success obtaining positive and negative emotion ratings during an affective situation (Fredrickson & Levenson, submitted). Dial rating during an interaction, however, may create a demand on subjects to monitor their affect that is disruptive to emotional experience. Also, this procedure cannot be used to obtain information about the possible occurrence of each of a number of specific emotions.

To the best of our knowledge, no research has examined whether coherence between facial expression and self-report is discernible on a moment-by-moment basis. In the present research, we chose to deal with the temporal and memory limitations of self-report of emotion described above, by having subjects rate their emotions on a momentary basis and by providing them with a cue to aid them in remembering which emotions they had felt. Additionally, we examined momentary changes in facial expressions of emotion rather than aggregated facial measures, and then looked at the extent to which specific expressions and reports varied together.

Overview of Empirical Strategy

We showed subjects brief, emotionally evocative films, and asked them to report on their emotions using a new, moment-by-moment reporting procedure. This new procedure allows subjects to rate the extent to which

they experienced each of several categories of emotion for many locations over a time interval during which a number of emotions may occur, without interrupting the flow of emotional experience. In this method, called *cued-review*, subjects see a replay of the stimulus film and provide emotion reports at points when they remember having felt an emotion during their initial viewing. The cued-review technique provides reports of momentary changes in the subjective experience of emotion for a time period during which more than one emotion may have occurred. Thus, although cued-review is not a continuous measure of subjective experience, it may allow us to study coherence between facial expressions and self-reports of emotions that occur at specific moments during a stimulus period.

We expect that reports of emotion at specific moments should be sensitive to the ebb and flow of subjective experience—variations that may correspond to fluctuations in spontaneous facial expression of emotion. With cued-review we can determine whether there is congruence between reports of experience and facial expressions of emotion at specific points in a time.

Our first set of hypotheses postulate co-occurrence at specific points in time between facial expression and self-report of emotions. First, we will examine co-occurrence at specific moments on a between-subjects basis:

Hypothesis 1a. For film locations where there is a high frequency of facial expression of emotion across subjects, there should also be a high frequency of self-report of emotion.

Our next hypothesis posits that correspondences that appear on a between-subjects basis also occur within-subjects:

Hypothesis 1b. Within-subjects, self-reports of emotion and facial expression of emotion should occur at the same locations during a stimulus period.

It is also possible that the extent to which we observe coherence may depend on the intensity of the event. There are two possible explanations for intensity-dependent coherence in our study: as self-reports of momentary occurrences of emotion are obtained after the initial stimulus period, subjects may remember only their more intense experiences even with the aid of a visual cue; or it could be that before response systems will act in unison, emotional events may have to reach a minimum intensity level, as has been suggested recently by other emotion researchers (Davidson, 1992; Tassinary & Cacioppo, 1992). Co-occurrence between expression and report was examined for each subject's most intense versus least

intense facial expression and again for each subject's most intense versus least intense self-report locations:

Hypothesis 1c. People should be more likely to give self-reports of emotions during their most intense facial expressions of emotion than during their least intense facial expressions;

Hypothesis 1d. People should be more likely to show facial expressions of emotions during their most intense self-reports of emotion than during their least intense self-reports of emotion.

The most specific level at which to evaluate coherence is that of categorical agreement—the extent to which co-occurring facial expression and self-report agree as to the type of emotional response. The second set of hypotheses examines coherence at the level of categorical agreement, in order to answer questions such as this one: If a person shows a facial expression of fear did he or she also remember experiencing fear? We expect that this should be true generally, i.e. for all emotional expressions and reports (hypothesis 2a), and more specifically for more intense emotional events. Hypothesis 2b defines intensity on a between-subjects basis (in terms of more emotionally salient portions of the film), whereas 2c and 2d define intensity on a within-subjects basis (in terms of individuals' most intense facial expressions and self-reports, respectively).

Hypothesis 2a. When facial expressions of emotion and self-reports of emotion co-occur in time for any given subject, they should agree as to the type of emotion that occurred;

Hypothesis 2b. For film locations where there is a high incidence of facial expression and self-report of emotion across subjects, those subjects who have facial expressions and self-reports of emotion at these locations will demonstrate categorical agreement between these two measures;

Hypothesis 2c. Where people show their most intense facial expressions of emotion, they should also report having experienced the same type of emotion;

Hypothesis 2d. Where people provide their most intense reports of emotional experience, they should also show facial expressions of the same type of emotion.

METHOD

Subjects

Twenty female students were recruited from universities in the San Francisco area (Age, M = 22.6 years, SD = 4.37). They were paid $10 for the one hour of experimental time.

Design

All conditions were run on a within-subjects basis. Each subject viewed several film segments and rated her emotions immediately following each film using two methods: cued-review (CUE) and retrospective report (RETRO).[1] To control for order effects, the presentation sequence of the experimental films was varied between-subjects by a method of incomplete counterbalancing in which each film appeared once in each possible ordinal position. Whether RETRO or CUE ratings were obtained first for each of the experimental films varied by a method of incomplete counterbalancing. There were four possible orders for whether RETRO or CUE came first, five subjects were assigned randomly to each of these orders.

Procedure

Subjects sat at a table facing a 19 inch, colour CRT monitor. On another cart to the right of the large monitor, was a 5 inch, black and white television positioned above a VCR. The lens of the video camera (which was mounted in the adjacent room) was hidden in the glass cabinets behind the monitor. The experimenter sat in another room for the entire experiment following the initial instruction and practice periods, and was able to see the subject's face on a video monitor. The experimenter controlled the playing of the stimulus tapes during the initial viewing (emotion-induction). The tapes used for playback during cued-review were controlled by the subject and played on the VCR in the experimental room.

Emotion Induction Films. Subjects viewed six film segments, four of which had been edited to a length of 1:14–1:18 minutes. These four film

[1] Restrospective emotions reports in which subjects provided a single report on emotion for each film were collected to examine coherence at a summary level (as has been done in previous research), and to determine which had a stronger relationship with aggregated facial behaviour: retrospective emotion reports or aggregated CUE reports. As the focus of this paper is on momentary correspondences and the aggregate analyses deal with a separate issue, those findings will be presented elsewhere.

clips, which were the experimental induction films, had been selected for their ability to elicit primarily disgust and secondarily fear and other negative emotions, in previous research.[2] Disgust was chosen for the primary emotion, because it has distinct facial expressions that can be reliably elicited in laboratory exposure to films (Ekman et al., 1980). We included only two of the four films for analysis: (RAT), which depicted a scene where a rat approaches a sleeping man, and eventually enters his mouth;[3] and (AMP), which contained scenes of burn patients and an amputation procedure from a medical training film (similar versions of AMP were originally used for emotion induction by Ekman & Friesen, 1974). The other two film segments were omitted from analysis, because they were both scenes from popular horror or science fiction movies that many of the subjects reported having seen already. We were concerned that remembered affect from previous viewings of these films or knowledge of socially consensual reactions to these films could influence affect ratings and reactions in the laboratory.

A fifth and sixth film segment, both of which were 1 to 2 minutes in length and which elicit primarily happiness (a fairy tale ending to a romantic film and a scene of a puppy playing with a flower) were not part of the experimental manipulation, but always appeared as the last segments for each subject to alleviate any unpleasant effects of the first four films (subjects gave ratings for both of these segments as well).

All subjects were tested individually. The experimenter told the subject that she would be seeing a series of short film clips, some of which would be unpleasant, and that we were interested in her emotional reactions to them. The subject then had a practice trial in which she received instructions for and practised each rating technique. The experimenter was in a separate room during the practice trial. The film stimulus used for the practice trial was selected for its ability to elicit mild enjoyment (scenes of monkeys and gorillas playing in a zoo, from Ekman & Friesen, 1974). The order in which each subject performed the rating procedures in the practice trial was counterbalanced across subjects.

Rating Techniques. We derived the cued-review procedure from Kagan et al.'s (1965) interpersonal process recall technique and Levenson and Gottman's (1983, 1985) video-recall technique, but it differs from these methods in that the replayed videotape is not a recording of the

---

[2] Previous studies of students' ratings of emotion elicited by some of these films—conducted by James Gross, University of California, Berkeley—were helpful in our selection of stimulus materials.

[3] The RAT film was originally used for emotion induction by Richard Davidson, of the University of Wisconsin at Madison.

subject interacting with a partner. Instead, it is the film that was used for emotion induction. In cued-review, subjects are instructed to stop the video replay at the points where they remember having felt an emotion during their first viewing of the stimulus film. They then use an emotion report form to rate what they remember feeling at that precise location in the film. After completing a given rating, subjects then start the video again and stop the playback whenever they remember having felt a *change in emotion* (in either degree or type) during the first viewing. Subjects then complete another form for that location. Thus, subjects go back through the entire stimulus segment and provide a series of emotion reports about what they remember feeling during the initial viewing period. We call this technique cued-review, because it was designed so that subjects could use the visual cue of the stimulus film to help them remember what they felt the first time they saw the segment. In this study, the experimenter independently recorded the locations of each subject's stopping points from the adjacent room.

For both the CUE and RETRO ratings, each rating form contained eight emotion terms: anger, contempt, disgust, embarrassment, fear, happiness, sadness, and surprise, each of which was rated on a scale from 0 to 8, where "0" represented "not the slightest bit" of the emotion and "8" represented "the most I have ever felt in my life". This method for anchoring the scale in terms of life experience was developed by Ekman and first reported in Levenson et al. (1990).

In the RETRO condition, subjects were told to report what they remembered feeling when they viewed the film on one form, called the "post-film emotion form". Subjects were asked to rate the extent to which they felt each of the eight emotion terms over the entire film period using the 0 to 8 scale. At each stopping point in the CUE procedure, the subjects completed emotion forms called "review-film emotion forms". The review-film emotion forms included the same terms and scales as those in the RETRO condition. The only difference was that subjects rated what they remembered feeling in the film during the initial viewing *at the point in the film where they had stopped the tape* during the CUE procedure.

In the CUE condition, the video was played back on a small, black and white monitor and with no sound. This was done to reduce the emotion eliciting properties of the stimulus film on re-viewing, because we wanted to minimise the chances of subjects reporting on re-experienced rather than remembered affect in the CUE condition. On debriefing, no subject reported re-experiencing affect during cued-review.

In the practice period, each subject received instructions for the rating techniques. Then they had two practice trials, as described above. Once the subject and experimenter were satisfied that the subject understood the procedures, the experimenter left the room and did not communicate

with the subject until all films had been shown and all ratings had been made. For all subjects, each film segment first appeared on the large colour monitor. Then instructions appeared on the screen that told subjects which rating procedure to use.

Each of the following intervals was 30 seconds in length: the interval between the end of the film segment and video replay (for conditions when the CUE ratings were made before the RETRO), the interval between the end of each film segment and RETRO ratings when that condition was first, the interval between the end of one rating procedure and the beginning of the other, and the interval between the end of a rating procedure and the start of the next film.

Video Recordings

Spontaneous facial behaviour exhibited during film viewing was videotaped via concealed camera. Not one subject guessed that she had been videotaped. After the experiment was completed, subjects were informed of the camera, and written consent was requested for the use of the recordings for scientific purposes. Subjects were told that if they did not want their tapes analysed, we would destroy the tapes and they would still receive full payment for their time. No subject denied permission for the use of her videotape.

Facial Data

Facial behaviour was scored using Ekman and Friesen's (1978) Facial Action Coding System (FACS). FACS dissects all observable facial movement into 44 visually distinguishable and anatomically separate units of muscular action. We scored every facial event that occurred in terms of the action units that singly or in combination with other action units produced it. We then grouped combinations of action units into specific emotion and nonemotion categories on the basis of an empirically and theoretically derived computer database of expression patterns for use in the data analyses below.

Intercoder reliability for FACS coding has been well established across several laboratories (Ekman & Friesen, 1976; Ekman, Friesen, & Simon, 1985; Fox & Davidson, 1988; Krause, Steimer, Sanger-Alt, & Wagner, 1989; Steiner, 1986). The two coders for this study each had more than one year's experience as a FACS coder, and their reliability had been established against a standard criterion (Ekman and Friesen's own scoring) prior to scoring. Interrater agreement for this sample was calculated by obtaining a ratio of the number of agreements between the coders divided by the total number of agreements and disagreements. The agreement ratio between coders was 0.81.

In sum, from each subject there were records of facial behaviour exhibited during the film viewing period (which were later FACS coded), a single retrospective report of her subjective experience during each film, and multiple CUE reports on the same experience.

RESULTS

Was there sufficient incidence of facial expression and CUE reports to be able to determine whether they co-occurred?

In order to test our hypotheses, it was necessary to establish that there were multiple reports and multiple expressions for each film, and that each film provided a comparable amount of opportunities to test correspondence on a momentary basis. Overall, subjects stopped the film during the cued-review condition an average of 5.51 times (SD = 2.36). The median and modal number of stopping points were five. The mean number of stopping points did not differ significantly between films (paired $t(19) = 1.89$, $P = 0.074$), but on average, subjects provided more CUE reports for the AMP film than the RAT (AMP $M = 5.95$, RAT $M = 5.10$). The mean number of facial expressions of emotion did not differ significantly between films (RAT $M = 4.50$, AMP $M = 6.75$, paired $t(19) = -1.74$, $P = 0.098$), but there tended to be more expression in response to the amputation film. It appears that a sufficient amount of expression and reporting occurred for both films to proceed with analyses of specific emotions at several points in time.

All analyses presented below use all emotions that subjects exhibited. Coherence was always evaluated idiographically, in terms of whether a subject's facial expression and self-report of emotion were congruent. Disgust was the emotion exhibited most often by most of the subjects, but other emotions occurred. For each of the films, facial expressions of sadness occurred more than 17% of the time, fear 10% of the time, and expressions of happiness, contempt, anger, and nonspecific negative emotion each between 3% and 5% of the time. Across all facial actions scored on a 5-point intensity scale, the mean level of intensity for RAT = 2.5, SD = 1.9; for AMP, $M = 3.34$, SD = 5.4.

For the CUE reports for RAT, the majority of the reports listed fear or disgust as the most intense or "peak" emotions, with only a few occurrences of other emotions, such as surprise and sadness, as the most intense reported emotion. For AMP, disgust was the overwhelmingly dominant response in CUE reporting, with only a few reports listing surprise, fear, anger, and sadness as the peak emotion. For RAT, the mean intensity for CUE ratings of disgust was 3.02, SD = 1.34; fear $M = 3.19$, SD = 1.64. For AMP CUE reports, disgust $M = 4.69$, SD = 1.81; fear $M = 1.62$, SD = 1.67.

The remaining analyses are organised according to a series of five analytical questions relevant to the specific hypotheses. Questions 1 and 2 are relevant to Hypotheses 1a–1c. Questions 3–5 concern Hypotheses 2a–2d.

Question 1: *Did facial expressions of emotion and self-reports co-occur in time?*

The first set of hypotheses addressed coherence at the temporal level, and posited that facial expressions and CUE reports of emotion would co-occur in time. Hypothesis 1a predicted that for film locations where there was a high frequency of facial expression across subjects, there would also be a high frequency of self-report of emotion. In order to test this hypothesis, we first needed to determine whether certain locations of the film were more active in terms of expression than others.

Determining Facially Active Locations. For each film, there were certain locations where more than half of the subjects showed facial expressions of emotion. We conducted binomial tests to determine whether the proportion of subjects who exhibited a facial expression of emotion at these "active" locations differed from what would have been expected to occur by chance at any given film location. Chance probabilities were determined by sampling total film time at moving 4sec windows, and then calculating the probability of an emotion expression occurring across subjects at each interval (the number of subjects showing a facial expression of emotion/total number of subjects). We then averaged these multiple probabilities to yield a chance or "expected" probability of an emotion expression occurring at any given 4sec interval in each film. We chose 4sec for the interval length on the basis of previous research that has shown that most facial expressions of emotion last between 0.5sec through 4sec (Ekman, 1984). Chance probabilities were calculated separately for each film. The proportion of subjects showing an emotional expression during each active interval was then compared to the chance level.

TABLE 1
Observed and Chance Probabilities of Subjects Showing a Facial Expression of Emotion at Facially Active Locations

Film	*Proportion Observed*	*Proportion Expected by Chance*	*P*
RAT	0.75	0.20	< 0.001
AMP			
Location 1	0.50	0.25	< 0.01
Location 2	0.70	0.25	< 0.001

Note: Probabilities based on binomial tests, $N = 20$.

Indeed, there were locations within each film where most subjects showed significantly more facial expressions of emotion than at any given 4sec interval of film time. For the RAT film, there was one such "facially active location": for AMP there were two.

Table 1 shows that for each of the facially active film locations, the proportion of subjects who showed a facial expression of emotion was significantly greater than the proportion expected by chance to occur at any given 4sec window of film time. This finding held for both of the films, and for both candidate locations within one of the films (AMP). No other locations in either film differed substantially from chance.

Now that we have established that the film locations of peak facial expression of emotion activity were significantly more active than any other given 4sec window of film time—for both films—we can test Hypothesis 1a: There should be a high incidence of self-report of emotion at film locations where facial expressions of emotion were more likely to have occurred. We conducted binomial tests to determine whether the proportion of subjects who provided a CUE report of emotion at the facially active locations was significantly greater than would have been predicted by chance. For each film, chance probabilities were calculated by sampling moving 4sec windows of film time, and then calculating the probability of a CUE report occurring across subjects at each interval (the number of subjects who gave a CUE report/total number of subjects). We then took the mean of these multiple probabilities to yield a chance or "expected" probability of a CUE report occurring at any given 4sec interval of film time.

We compared the observed proportions of subjects who provided CUE reports at each facially active location with the chance values. Table 2 shows that for each of the facially active locations, significantly more subjects provided a CUE report than would have been expected by chance. This finding held for all facially active locations across both films.

The fact that significantly more subjects provided CUE reports at facially active locations than would have been expected by chance supports

TABLE 2
Observed and Chance Probabilities of Subjects Providing CUE Reports of Emotion at Facially Active Locations

Film	*Proportion Observed*	*Proportion Expected by Chance*	*P*
RAT	0.95	0.29	< 0.001
AMP			
Location 1	0.80	0.25	< 0.001
Location 2	0.95	0.25	< 0.001

Note: Probabilities based on binomial tests, $N = 20$.

Hypothesis 1a. Even more compelling evidence for the coherence question is the fact that the 4sec windows of film time that were peak locations of emotional facial expression activity (i.e. the facially active locations) were also the locations of highest CUE reporting for both films.

Hypothesis 1b framed the question of co-occurrence on a within-subjects basis, and predicted that for individual subjects facial expressions of emotion and CUE reports of emotion would tend to occur at the same points in film time. We included facial expressions and CUE reports of all emotions in this analysis. Although the films were selected for their ability to arouse primarily disgust and fear, occasionally they evoked other emotions. As Hypothesis 1b is a question of general co-occurrence that is not asked at the level of specific types of emotions, CUE reports and facial expressions of anger, contempt, disgust, fear, happiness, sadness, and surprise were included in this analysis.

In order to test Hypothesis 1b, we calculated a co-occurrence quotient for each subject: The total number of times a CUE report coincided with a facial expression of emotion/total number of facial expressions of emotion. A CUE report was considered to be "coincident" if it overlapped with a facial expression of emotion within a 4sec window of film time (i.e. plus or minus 2sec). For each film we took the mean of the co-occurrence quotients across subjects. We then compared this mean co-occurrence quotient to a hypothesised chance quotient. The chance values were defined as the probability of a CUE report occurring at any given 4sec interval of film time. They were calculated as follows: For both RAT and AMP, the modal number of CUE points equalled 5. Multiplying this by the 4sec sampling window yields 20sec of film time allocated for CUE reports. Total film time for RAT was 78sec, thus 20/78 or 0.26 of film time was allocated to CUE reporting. For AMP, total film time was 74sec, so 20/74 or 0.27 of film time was allocated to CUE reporting. For ease of analysis, we rounded up both numbers to yield a chance reporting quotient of 0.30 for both films. Rounding up is a conservative strategy and works against our hypothesis.

To test Hypothesis 1b, we then compared the observed mean co-occurrence quotients with the chance quotients in one sample *t*-tests. For RAT, the mean co-occurrence quotient across subjects was 0.46 (SD = 0.28). This was significantly greater than the chance value of 0.30 ($t(16) = 2.305$, $P < 0.02$). For AMP, the observed co-occurrence quotient was 0.47 (SD = 0.22), which was also significantly greater than the chance value of 0.30 ($t(18) = 3.409$, $P < 0.01$).

The question of co-occurrence can be asked from the point of view of the CUE reports as well: To determine whether subjects showed facial expressions at points in each film where CUE reports occurred. When we conducted the same analyses using CUE report locations as the starting

point and looked for co-occurring facial expressions, we obtained nearly identical results; therefore, the details of those analyses are omitted.

The above results support Hypothesis 1b, showing that on the average, co-occurrence of any type of emotional facial expression and self-report was greater than would have been expected by chance, whether the question is framed from the perspective of facial expression or CUE report of emotion. The observed co-occurrence quotients, however, did not indicate co-occurrence for all CUE reports and facial expressions of emotion. On the average—for both films—CUE reports occurred during the same film locations as facial expressions of emotion about half of the time. This raises questions about the conditions that influenced whether or not particular expressions and reports co-occurred.

Question 2: *Did co-occurrence between expression and report vary as a function of intensity of emotion?*

Our next two hypotheses predicted that within-subjects, co-occurrence may be more likely for more intense emotional events. Hypotheses 1c and 1d posited that people would be more likely to give CUE reports at locations of their most intense facial expressions of emotion than of their least intense facial expressions and vice versa: People would be more likely to show emotional expressions at the same locations in film time as their most intense CUE than their least intense CUE reports. To test Hypothesis 1c, we determined the number of subjects who gave a CUE report within a 4sec interval of their *most* intense facial expression of emotion, and how many provided a CUE report at the same locations as their *least* intense facial expressions of emotion. We then conducted McNemar tests of change in related samples (Siegel, 1956) to determine whether co-occurrence was more likely to occur during most intense emotional expressions versus least intense emotional expressions.

Determining Intensity of Facial Expressions of Emotion. In FACS, action units (AUs) that are critical to distinguish among certain emotional expressions are scored on a 5-point intensity scale (Friesen & Ekman, 1992). To determine the intensity of an expression, we summed the intensity scores for each critical AU in the expression to yield a total intensity score. For each subject (for each film) we designated the emotional expression with the highest total intensity score as the most intense facial expression of emotion. If there was a tie on total intensity score, we chose the expression with the longest duration. We designated the expression with the lowest sum intensity score as the least intense facial expression of emotion. If there was a tie for lowest total intensity score, we chose the expression with the shortest duration.

For RAT, 16 of the 20 subjects gave CUE reports at the same points in film time as their most intense facial expressions of emotion; only four did so during their least intense expression. The McNemar test showed that this was a statistically significant difference ($\chi^2 = 10$, $df = 1$, $P < 0.01$). For AMP, 19 subjects gave CUE reports during their most intense facial expressions of emotion, but only seven did so during their least intense emotional expressions. The difference between these proportions was also significant ($\chi^2 = 10$, $df = 1$, $P < 0.01$). These results support Hypothesis 1c.

Hypothesis 1d examined the same question framed from the standpoint of most intense CUE reports, predicting that people would show a facial expression of emotion at the same locations as where they gave their most intense reports of emotional experience.

Determining Intensity of CUE Reports of Emotion. For each report, subjects rated each emotion term on a 9-point intensity scale. We summed the intensity scores for all emotion terms on each report. Then, for each subject, we designated the report with the highest total intensity as the most intense CUE report, and the one with the lowest total intensity as least intense CUE report. Cases of tie were settled by random selection.

Again we conducted McNemar tests. For the RAT film, 16 of the 20 subjects showed a facial expression at the same location as their most intense emotion reports. Only seven subjects showed emotional expressions during their least intense emotion reports. This difference was significant ($\chi^2 = 10$, $df = 1$, $P < 0.01$). For AMP, however, there were no significant differences in the number of subjects who showed facial expressions of emotion during most intense CUE reports versus least intense CUE reports (16 vs. 13, respectively, $\chi^2 = 1.29$, $df = 1$, $P > 0.20$). The results from three tests of Hypotheses 1c and 1d support the intensity hypothesis on a within-subject basis. The one exception will be considered in the discussion section.

Question 3: *Was there categorical agreement between co-occurring facial expressions and self-reports?*

The second set of hypotheses examined coherence in terms of categorical agreement on the type of emotion between co-occurring facial expression and self-report. Hypothesis 2a predicted agreement between expression and CUE report. To test this hypothesis we conducted analyses similar to those for Hypothesis 1b, only this time we compared categorical *agreement* coefficients to the hypothesised chance values. We calculated agreement or "matching" quotients for each subject defined as: The number of times a CUE report that coincided with a facial expression of emotion agreed or

matched as to the type of emotion that occurred/total number of facial expression – CUE report co-occurrences. Matching was defined in terms of whether emotional category of facial expression was the same as the category of the emotion rated at highest intensity on the CUE report.

Because the distributions of matching quotients across subjects were extremely non-normal, one sample Wilcoxon signed ranks tests were conducted to compare observed quotients to hypothesised population chance values (Shott, 1990). We used the same chance values that were used in the test of Hypothesis 1b, which reflected the probability of a CUE report occurring at any given 4sec interval of film time. We reasoned that these chance values—as expected values of co-occurrence and not necessarily matching between expression and report—would necessarily be high estimates of matching. Thus, our tests of difference from this chance value would be very conservative. For RAT, the observed median matching quotient was 0.42. This was not significantly different from the chance level of 0.30 ($T(16) = 83$, $P > 0.10$).[4] For AMP, the sample median quotient was 0.50, which was significantly greater than the chance value of 0.30 ($T(19) = 149$, $P < 0.025$). Thus, only results from the AMP film support the hypothesis of overall matching between facial expressions of emotion and co-occurring CUE reports posited in Hypothesis 2a. When we evaluated categorical matching for facial expressions that occurred at the same locations as CUE reports, we obtained results identical to those obtained from the above analyses.

Co-occurring reports and facial expression agreed on category of emotion only for the AMP film. These analyses were conducted across each film, for all locations at which CUE reports and facial expressions co-occurred. Would there be greater evidence of agreement at points in each film that were more emotionally intense? Question 4 examines this by looking at categorical agreement at points during each film at which most subjects showed emotional responses (thereby defining intensity on a between-subjects basis). Question 5 looks at whether categorical agreement is greater at the most intense points during each film at each subject's location of greatest expressivity or report (thereby defining intensity on a within-subjects basis).

Question 4: *Was there categorical agreement between co-occurring facial expression and self-reports at emotionally active locations?*

[4] T = the sum of the positive signed ranks, and the number in parentheses refers to the total number of signed ranks.

TABLE 3
Observed and Chance Conditional Probabilities of Subjects Matching on Emotion Category given that they had shown a Co-occurrence of Facial Expression and CUE Report at Emotionally Active Locations

Film	*Proportion Observed*	*Proportion Expected by Chance*	*P*
RAT	0.71 (n = 14)	0.20	< 0.001
AMP			
Location 1	0.90 (n = 10)	0.25	< 0.001
Location 2	0.83 (n = 12)	0.25	< 0.001

Note: Probabilities based on binomial tests.

Hypothesis 2b predicted categorical agreement between co-occurring facial expression and CUE report of emotion at the emotionally active locations (film locations characterised by a high degree of expression and report). A new set of chance conditional probabilities were computed. For moving 4sec intervals of film time we determined the probability of a match on type of emotion given a co-occurrence between CUE and facial expression, and then took the mean of these to yield chance probabilities of co-occurrence at any given 4sec interval of film time. We then conducted binomial tests to compare the observed and chance probabilities of a subject matching on emotion category given temporal correspondence of facial expression and CUE report.

As seen in Table 3, the proportion of the subjects whose expression and report matched on category for the RAT film—given that they had demonstrated co-occurrence between facial expression and CUE report of emotion at the emotionally active location—was significantly greater than would have been predicted by chance. Similar results occurred for both of the emotionally active locations in the AMP film, where significantly more of the subjects who showed temporal linkage between facial expression of emotion and CUE report at active locations actually matched between measures on emotion category. These findings support Hypothesis 2b.

Question 5: *Was categorical matching a function of intensity?*

Hypothesis 2c predicted that where people show their most intense facial expression of emotion, they will also report experiencing the same type of emotion. In order to test this hypothesis, we calculated the proportion of subjects whose CUE report and facial expression matched on category *given* they had demonstrated co-occurrence between the two measures for the most intense facial expressions of emotion and we compared these to chance values via binomial tests. For RAT, 12 of the 16 subjects who had CUE reports concomitantly with most intense facial expression matched on category. For AMP, 13 of the 19 subjects who demonstrated co-occurrence

between most intense facial expression and CUE reporting matched. Each of these proportions was compared to a chance value of 0.30 (derived in the test of Hypothesis 1b) in a binomial test. Both observed proportions were significantly greater than chance (RAT, $P < 0.001$; AMP, $P < 0.001$). We then looked at matching for the least intense facial expression of emotion.[5] For RAT, of the four subjects who demonstrated co-occurrence at their least intense facial expression of emotion, only one matched on category. This proportion was not different from chance ($P > 0.40$). For AMP, of the seven subjects who gave a CUE report at the same location as their least intense facial expression, only two matched on emotion category, which did not differ significantly from chance ($P > 0.30$). These results support Hypothesis 2c.

Hypothesis 2d framed the within-subjects categorical agreement question from the standpoint of most intense CUE report. For each film, we calculated the proportion of subjects who—given they showed a facial expression of emotion at the same film locations as their most intense CUE report—matched on emotion category. This was then compared to chance values. Chance values for these analyses were based on the average number of times subjects showed facial expressions of emotion per film. With one exception (to be discussed later) the result obtained from these analyses were the same as those from the analyses of Hypothesis 2c. For both films, of the subjects who showed facial expressions of emotion at same film locations as their most intense CUE reports (10 out of 16 for both films) a significantly greater than chance proportion matched on emotion category ($P < 0.01$). Again, as in the tests of Hypothesis 2c, this did not hold for least intense CUE reports. Only 1 of the 7 subjects who showed co-occurrence between most intense report and facial expression for RAT matched on category. This was not different from chance ($P > 0.30$). These results support Hypothesis 2d. The one exception occurred for AMP, in which 8 of the 13 subjects for whom co-occurrence for least intense CUE report matched on category. This is different from the chance level of 0.30 ($P < 0.05$). This exception will be considered in the Discussion section.

Although the analyses on categorical agreement included facial expressions and self-reports of any type of emotion, it is possible that the high levels of categorical agreement between co-occurring facial expressions and CUE reports described above were a necessary product of the predominance of one type of emotion being shown facially or reported as the peak emotion within and across subjects. If the films elicited primarily one type of

[5] We could not look at the difference in proportions of subjects who matched in the most intense versus least intense condition using the McNemar test, due to the restriction of our sample to those who showed co-occurrence.

emotion, then any facial expression and self-report that occurred together would have to agree on category. We have already noted that both films elicited primarily disgust. Have we established evidence of specificity, or was categorical agreement just a necessary byproduct of temporally linked expression and report during films that evoked primarily disgust?

The ideal test of this question would be to determine whether categorical agreement of facial expression and CUE report held for nondisgust expressions and reports. Unfortunately, while other emotions occurred in both films, they occurred too infrequently to allow for adequate analysis of nondisgust emotions as a separate group. Also, when other emotions did occur, they were never the most intense instances of expression and report. The confounding of intensity of response with type of emotion would also complicate analysis. Thus, we chose an alternate means of determining whether categorical agreement between expression and report was a necessary byproduct of expressions and reports overlapping in time for films that elicited primarily disgust.

For each subject's most intense facial expression of disgust, we determined whether or not she reported disgust or another negative emotion as the most intense emotion on the corresponding CUE report. We reasoned that if a variety of negative emotions were reported by subjects during their most intense disgust facial expressions—but disgust was reported as most intense significantly more often than other negative emotions—then we would have evidence that categorical agreement was not merely a result of co-occurrence. We conducted χ^2 goodness-of-fit tests to determine whether the distribution of subjects who reported disgust and nondisgust negative emotions as the most intense emotion on the CUE report that corresponded with the most intense facial expression of disgust differed from chance. Probabilities for the chance proportions were calculated as follows: We determined the proportion of each subject's negative emotion reports in which disgust was reported as the most intense or "peak" emotion, and then we took the mean of these proportions across subjects. By this method, for RAT chance distribution of negative emotion report at any film location was: 35% of subjects reported disgust as the peak emotion, and 65% reported other negative emotions as peak. The observed distribution for RAT was: 75% of the subjects who gave a report at the same location as most intense facial expression reported disgust as the peak emotion, and 25% reported experiencing some other negative emotion as peak. This distribution was significantly different from the one expected to occur by chance ($\chi^2 = 8.44$, $df = 1$, $P < 0.005$). For AMP, the chance distribution at any film locations was: 81% of the subjects reported disgust as the most intense, and 19% of the subjects reporting other negative emotions as the most intense. In fact, 80% of the subjects who gave a report at the same location as their most intense facial expression reported

disgust as the peak emotion, and 20% reported experiencing some other negative emotion. Not surprisingly, this distribution was not different from the distribution expected by chance for AMP ($\chi^2 = 0.01$, $df = 1$, $P > 0.50$). The same results hold when the question is framed with most intense disgust CUE report as the starting point. Thus, for AMP—in which disgust was the emotion reported and shown facially most of the time—categorical agreement was a necessary consequent of co-occurring expression and report. For RAT, however, as more types of emotions were reported, categorical agreement indicates specificity in coherent emotional responses.

DISCUSSION

Our results provide the first evidence that there is coherence between facial expression and self-report of emotion at specific moments. For emotions that occurred during a brief episode, we obtained evidence linking facial expression and self-report not only in time, but also in type. Thus, we have extended previous findings of correlations between single indices of expression and report of emotion over a period of time (Ekman et al., 1980), and have shown that correspondences exist at specific moments, in particular for intense emotional events.

We found evidence for coherence at several levels of analysis. When we analysed coherence at the level of co-occurrence between facial expression and self-report of emotion at several locations in film time, we found that facial expression and self-report coincided more often than would have been expected by chance. This result held between- and within-subjects, and whether we framed the question with either facial expression or self-report as the reference from which to evaluate co-occurrence. Overall, the level of within-subjects co-occurrence between self-report and facial expression of emotion was about 50% for each film. This percentage, though significantly greater than the chance values, still suggests quite a bit of nonoverlap. This moderate amount of overlap would be expected, however, when one considers that this value is an index of general co-occurrence between report and facial expression of all intensities and categories. When we asked the question in terms of coherence between expression and report of the same types of emotions, the degree of correspondence improved.

As predicted, we found that co-occurrence was optimal for more intense emotional events, whether we based intensity on facial expression or self-report. The idea that intense emotional responses are more likely than less intense responses to involve linkage between expression and report was substantiated by the fact that in nearly all instances, significantly more subjects demonstrated co-occurrence for their most intense responses than for their least intense responses. The exception to this was that for AMP,

both the most intense CUE and least intense CUE showed a high incidence of co-occurrence between face and report. One possible explanation for this is that the AMP film, which elicited more facial expressions of emotion and CUE reports than the RAT film in this sample (though not significantly so), may have elicited more intense emotion overall. Specifically, the least intense CUE reports for AMP might have been more intense than the least intense CUE reports for RAT. A *post hoc* *t*-test revealed that this was indeed the case ($t(19) = 4.49$, $P < 0.001$).

Our finding that response cohesion may vary as a function of the intensity of the response is consistent with the view that emotions must reach a certain level of intensity before there can be congruence in response systems (Davidson, 1992; Tassinary & Cacioppo, 1992). Although our results are certainly suggestive of the existence of an intensity threshold for congruence in the entire emotional response, there are at least three reasons why this study is not definitive about this matter. First, the self-reports were not made online, but were obtained during a video replay of the stimulus film. This allows for the possibility that emotions had to reach a minimum level of intensity for their locations in film time to be recalled. Perhaps only the more intense emotions were remembered at the points in each film during which they were actually experienced. The less intense emotions may have been washed out in memory by the more intense ones.

Second, emotional experience may have to reach a minimum level of intensity to overcome problems of symbolic representation in verbal self-report. A person may be consciously experiencing an emotion that is manifest behaviourally and physiologically, but he or she may be unable to express the experience to others. Subjective experiences can be ineffable.

Third, emotions may have to be sufficiently intense to reach consciousness in the first place, let alone to be reported on. Other aspects of the emotion response—such as the connections between behaviour and physiological change—may not depend on intensity. Information from these systems need not enter consciousness to be measured. Thus, it is possible that the intensity effect is relevant to the entry of an emotion into conscious awareness, and not necessarily relevant to the coherence of other aspects of the emotional response. Physiology and behaviour may cohere at lower intensities of emotion, at which emotional experience is not yet perceived.

We cannot rule out the possiblity that memory problems or the problem of symbolic representation of experience accounted for our intensity-dependent findings nor can we disentangle these factors from the possibility that coherence in real-time depends on intensity. The third problem is more complicated. Intensity of an emotion at the behavioural and physiological levels may moderate the entry of the subjective system into

a coherent emotional response.[6] The effect of intensity on coherence is a complex question that can be asked at many levels: in terms of coherence within non-conscious responses systems (spontaneous behaviour and physiology), and in terms of coherence between conscious systems and non-conscious systems. The research presented in this paper is only suggestive of an intensity effect that should be explored further.

We obtained evidence for coherence at the most specific level—categorical agreement between concomitant facial expressions and self-reports of emotion. When we looked at film locations where most subjects showed facial expressions and gave CUE reports, a substantial and significant majority of the subjects demonstrated categorical agreement between facial expression and self-report of emotion. When we examined the question of agreement or "matching" between response measures for all emotion expressions and all CUE reports, we obtained evidence for a significant amount of overall categorical agreement between facial expressions that co-occurred with CUE reports in the AMP film only. This is consistent with our notion that the AMP film may elicit more intense emotion overall. When we separated most intense from least intense events for analysis, however, for both films and for both ways of framing the question (from the standpoint of facial expression or CUE report as most intense) there was significantly more matching than would have been predicted by chance for the most intense events only. For AMP only, this held for least intense CUE reports as well, which was consistent with the above findings on this film's intensity.

For both films, subjects' facial expressions and self-reports matched on category—not just valence—for most intense events. This is consistent with findings that relationships between facial expression and self-report of emotion discriminate among different positive and negative emotions. For example, Ekman et al. (1980) found that measures of disgust facial expressions were positively correlated with self-reports of disgust but not with reports of the negative emotions of anger or sadness. In terms of the specificity of positive emotion, smiles involving both raising of the lip corners and contraction of the muscles around the eyes (the Duchenne smile) have been shown to have a specific relationship with reported amusement over and above other positive emotions, such as excitement and happiness (Ekman et al., 1990). Thus, in the study of the relationship between facial expressions and self-report of emotion, it is crucial to obtain ratings on several categories of emotions. The dimensional view that emotional variability can be thoroughly explained by the domains positive and negative affect (Diener & Emmons, 1984; Watson & Tellegen, 1985)

[6] We thank an anonymous reviewer for suggesting this.

may apply to general mood states, but our findings suggest that it is not adequate for the study of momentary emotional response.

One limitation of our findings was that for one of the films (AMP) categorical matching was a necessary consequence of co-occurring expression and report, because this film overwhelmingly elicited one type of emotion —disgust (80% of the emotional responses throughout AMP were disgust). That is, for a film that elicits only disgust expressions and reports, any co-occurring measures would by definition agree on category. For the RAT film, however, enough different emotions were reported to enable us to determine that categorial agreement between co-occurring expression and report was not merely a byproduct of co-occurrence. For RAT, even though subjects reported a variety of negative emotions during their most intense expressions of disgust, a significant majority of them remembered experiencing disgust as their most intense emotion rather than other negative emotions.

An important step in this research programme is to determine how well temporal linkage and agreement between facial expression and self-report generalise to other types of emotion as well as to other subject populations. Our criterion of coherence for the analyses between facial and reported emotion was primarily within-subject, on the basis of whether each subject reported in her CUE report the same emotion that she displayed facially at particular locations in film time. Even though there were some subjects who demonstrated coherence who were reporting and showing anger, fear, surprise, or sadness, most of the time the films elicited disgust. We limited our sample to college-aged women. We believe, however, that coherence between facial expression and self-report observed at specific moments should occur in other populations such as men, older adults, and children.

It may be suggested that the coherence domonstrated in this study was a particular to the stimulus situation; i.e. the viewing of disgust- and fear-eliciting films, and that it might not generalise to other emotional situations. It seems unlikely that humans would have evolved with emotional response systems that behave in one way to films but differently to other emotion elicitors. This possibility seems especially remote in the light of the fact that moderate levels of coherence among behavioural, physiological, and subjective measures have been demonstrated in the very different context of marital interaction research (Levenson & Gottman, 1983, 1985).

We demonstrated that measurement of facial expression and self-report of specific emotions on a momentary basis provides solid evidence of a coherent emotional response, especially for more intense emotional events. The source of this coherence remains unclear. It is possible that the observed coherence was internally driven, externally driven, or both. Our findings are consistent with Tomkins' (1962) notion of emotions as

correlated responses, and provide the strongest evidence to date that this is true for facial and subjective measures. Future research should strive to clarify the mechanisms responsible for coherence as well as demonstrate three-way coherence among subjective, behavioural, and physiological measures on a momentary basis.

Manuscript received 13 October 1992
Revised manuscript received 7 July 1993

REFERENCES

Block, J. (1957). Studies in the phenomenology of emotions. *Journal of Abnormal and Social Psychology, 54*, 358–363.

Brown, S., & Schwartz, G.E. (1980). Relationships between facial electromyography and subjective experience during affective imagery. *Biological Psychology, 11*, 49–62.

Buck, R. (1980). Nonverbal behavior and the theory of emotion: The facial feedback hypothesis. *Journal of Personality and Social Psychology, 38*, 811–824.

Cacioppo, J.T., Uchino, B.N., Crites, S.L., Snydersmith, M.A., Smith, G., Berntson, G.G., & Lang, P.J. (1992). Relationship between facial expressiveness and sympathetic activation in emotion: A critical review, with emphasis on modeling underlying mechanisms and individual differences. *Journal of Personality and Social Psychology, 62*, 110–128.

Campbell, D.T., Siegman, C., & Rees, M.B. (1967). Direction-of-wording effects in the relationships between scales. *Psychological Bulletin, 68*, 293–303.

Crowne, D.P., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Darwin, C. (1872/1965). *The expression of the emotions in man and animals.* University of Chicago Press.

Davidson, R.J. (1992). Prolegomenon to the structure of emotion: Gleanings from neuropsychology. *Cognition and Emotion, 6*, 245–268.

Davidson, R.J., Ekman, P., Saron, C.D., Senulis, J.A., & Friesen, W.V. (1990). Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I. *Journal of Personality and Social Psychology, 58*, 330–341.

Diener, E., & Emmons, R.A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47*, 1105–1117.

Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In J. Blacking (Ed.), *Anthropology of the body* (pp. 53–84). London: Academic Press.

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ekman, P. (1985). *Telling lies.* New York: Norton.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 161–168.

Ekman, P., Davidson, R.J., & Friesen, W.V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology, 58*, 342–353.

Ekman, P., & Friesen, W.V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29*, 288–298.

Ekman, P., & Friesen, W.V. (1976). Measuring facial movement. *Journal of Environmental Psychology and Nonverbal Behavior, 1*, 56–75.

Ekman, P., & Friesen, W.V. (1978). *The Facial Action Coding System.* Palo Alto: Consulting Psychologists Press.

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., & Simons, R.C. (1985). Is the startle reaction an emotion? *Journal of Personality and Social Psychology, 49*, 1416–1426.

Englander-Golden, P., Chang, H., Whitmore, M.R., & Dientsbier, R.A. (1980). Female sexual arousal and the menstrual cycle. *Journal of Human Stress, 6*, 42–48.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35*, 790–806.

Fox, N.A., & Davidson, R.J. (1988). Patterns of brain electrical activity during facial signs of emotion in 10-month old infants. *Developmental Psychology, 24*, 230–236.

Fredrickson, B.L., & Levenson, R.W. (Submitted). Positive emotion speeds recovery from negative emotional arousal.

Friesen, W.V., & Ekman, P. (1992). *Changes in FACS scoring.* Unpublished instructional document. University of California, San Francisco.

Jones, H.E. (1950). The study of patterns of emotional expression. In M.L. Reymert (Ed.), *Feelings and emotions* (pp. 161–168). New York: McGraw-Hill.

Kagan, N., Krathwohl, D.R., & Farquhar, W.W. (1965). *IPR—Interpersonal Process Recall: Simulated recall by videotape in studies of counseling and teaching-learning.* Michigan State University: East Lansing, MI.

Kleck, R.E., Vaughan, R.C., Cartwright-Smith, J., Vaughan, K.B., Colby, C.Z., & Lanzetta, J.T. (1976). Effects of being observed on expressive, subjective, and physiological responses to painful stimuli. *Journal of Personality and Social Psychology, 34*, 1211–1218.

Krause, R., Steimer, E., Sanger-Alt, C., & Wagner, G. (1989). Facial expression of schizophrenic patients and their interaction partners. *Psychiatry, 52*, 1–12.

Laird, J.D. (1974). Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology, 29*, 475–486.

Lanzetta, J.T., Cartwright-Smith, J., & Kleck, R.E. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology, 33*, 354–370.

Lanzetta, J.T., & Kleck, R.E. (1970). Encoding and decoding of nonverbal affect in humans. *Journal of Personality and Social Psychology, 16*, 12–19.

Levenson, R.W. (1988). Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In H.L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical implications* (pp. 17–42). New York: Wiley.

Levenson, R.W., Ekman, P., & Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*, 363–384.

Levenson, R.W., & Gottman, J.M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45*, 587–597.

Levenson, R.W., & Gottman, J.M. (1985). Physiological and affective predictors of change in relationship satisfaction. *Journal of Personality and Social Psychology, 49*, 85–94.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Orne, M.T. (1962). The social psychology of psychological experiments. *American Psychologist, 17*, 776–783.

Plutchik, R. (1962). *The emotions: Facts, theories, and a new model.* New York: Random House.

Reis, H.T., & Wheeler, L. (1991). Studying social interaction with the Rochester Interaction Record. In *Advances in experimental social psychology.* New York: Academic Press.

Shott, S. (1990). *Statistics for health professionals.* Philadelphia: Saunders.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Steiner, F. (1986). Differentiating smiles. In E. Baenninger-Huber & F. Steiner (Eds.), *FACS in psychotherapy research* (pp. 139–148). Zurich: Department of Clinical Psychology, University of Zurich.

Tassinary, L.G., & Cacioppo, J.T. (1992). Unobservable facial actions and emotion. *Psychological Science, 3,* 28–33.

Thomas, D.L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59,* 291–297.

Tomkins, S.S. (1962). *Affect, imagery, consciousness: Vol. 1.* New York: Springer.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.