Exploring the Strength of Association between the Components of Emotion Syndromes: The Case of Surprise

Rainer Reisenzein
University of Bielefeld, Germany

Publisher: Psychology Press, part of the Taylor & Francis Group
Issue: Volume 14, Number 1 / January 1, 2000
Pages: 1 - 38
URL: http://taylorandfrancis.metapress.com/link.asp?id=knlym5yf5r46n3wk
DOI: 10.1080/026999300378978

A new experimental paradigm involving a computerised quiz was used to examine, on an intra-individual level, the strength of association between four components of the surprise syndrome: cognitive (degree of prospectively estimated unexpectedness), experiential (the feeling of surprise), behavioural (degree of response delay on a parallel task), and expressive (the facial expression of surprise). It is argued that this paradigm, together with associated methods of data analysis, effectively controls for most method factors that could in previous studies have lowered the correlations among the components of emotion syndromes. It was found that (a) the components of the surprise syndrome were all positively correlated; (b) strong association existed only between the cognitive and the experiential component of surprise; (c) the coherence between syndrome components did not increase with increasing intensity of surprise; and (d) there was also only moderate coherence between the components of the facial expression of surprise (eyebrow raising, eye widening, mouth opening), although in this case, coherence tended to increase with intensity. Taken together, the findings support only a weakly probabilistic version of a behavioural syndrome view of surprise. However, the component correlations seem strong enough to support the existence of strong associations among a subset of the mental or central neurophysiological processes engaged in surprise.

INTRODUCTION

A widely held view in contemporary emotion psychology is that emotions, or at least a core subset of emotions, are organised patterns of more or less emotion-specific cognitive, experiential, behavioural (action-related), expressive, and physiological components. This assumption- which is frequently motivated by hypotheses about the evolutionary origin and biological function of emotions- is the common denominator of ``syndrome theories' ' of emotion, and it is shared by authors of otherwise fairly different theoretical persuasions (e.g. Averill, 1980; Ekman, 1992; Frijda, Ortony, Sonnemans, & Clore, 1992; Izard, 1977; Lazarus, 1991; Lazarus, Averill, & Opton, 1970; McDougall, 1908/1960; Plutchik, 1980; Roseman, 1996; Scherer, 1984; Smith & Scott, 1997; Tooby & Cosmides, 1990).[1] Considering the popularity of the syndrome view of emotions, it is surprising to learn that its empirical support is far from solid: Experimental

---

[1] Apart from ``syndrome view' ' (e.g. Averill, 1980; Lazarus et al., 1970) the theory of the nature of emotions under discussion also goes by names such as the ``componential' ' , ``multicomponent ' ' , or ``component process' ' view of emotions (e.g. Frijda et al., 1992; Scherer, 1984), the view that emotions are ``organised response patterns' ' , and probably still others. I prefer the term ``syndrome view' ' because it points to, but does not enforce, a probabilistic perspective, and because it is neutral with respect to the nature of the syndrome components (they can be mental, behavioural, or both).

studies of the degree of association among the components and subcomponents of presumed emotion syndromes (e.g. fear, happiness, anger, or surprise), particularly studies in which multiple syndrome components were measured, are comparatively rare; they exist only for few emotions; and they provide for the greater part at best limited support for the syndrome view. That is, these studies often found (a) non-significant or very weak, and occasionally even negative associations between syndrome components, as well as (b) context- (and participant-) dependence of these associations (for recent, partial reviews of the literature, see e.g. Meyer, Schützwohl, & Reisenzein, 1997a; Schmidt-Atzert, 1993). Although most of the more recent versions of syndrome theories of emotion are formulated probabilistically and do not even necessarily predict very high associations among (some) syndrome components, the near independence or even inverse relation of syndrome components suggested by these data is certainly problematic.

Among the first research to document a dissociation of presumed components and subcomponents of emotion syndromes were studies conducted by R.S. Lazarus and his co-workers in the early 1960s. These studies found that different physiological measures of ``psychological stress' ' induced by aversive films correlated only weakly with one another (e.g. Lazarus, Speisman, Mordkoff, & Davidson, 1962) as well as with subjective measures (e.g. Weinstein, Averill, Opton, & Lazarus, 1968), and that the correlations depended on context or participants (e.g. Lazarus, Tomita, Opton, & Kodoma, 1966). A number of subsequent studies that focused more narrowly on specific emotions confirmed these findings. For example, studies of fears and phobias often found low to missing correlations between the intensity of experienced fear assessed by self-reports, physiological arousal, and avoidance behaviour (e.g. Lang, 1988; Rachman & Hodgson, 1974). Similarly, studies of the association between self-report indicators of exhilaration on the one hand, and facial mirth expressions on the other hand, often reported low correlations (cf. McGhee, 1977; Ruch, 1990) as well as context-dependence (e.g. Hess, Banse, & Kappas, 1995). Analogous results were obtained in previous studies of the surprise syndrome, summarised later.

These findings could mean two things. First, they could mean that at least certain versions of syndrome theory (namely, those that assume relatively strong and context-independent associations among syndrome components), but perhaps even syndrome theories in general, are on the wrong track. This conclusion has indeed been drawn by several authors, who accordingly proposed to either abandon the concept of emotion syndromes (or even the concept of *emotion*) (e.g. Fridlund, 1994); or at least to drastically weaken the syndrome view, by assuming that the covarying response components form only a ``loosely coupled system' ' (e.g. Lang, 1988).

Second, the findings of missing, weak, or even negative associations between presumed emotion syndrome components could have been method artefacts (e.g. Cacioppo et al., 1992; Lazarus et al., 1962; Rosenberg & Ekman, 1994; Ruch, 1995; Weinstein et al., 1968). This hypothesis was initially the favoured response of emotion syndrome theorists to the empirical problem of lack of coherence among presumed syndrome components (cf. Weinstein et al., 1968) and it continues to be popular (e.g. Rosenberg & Ekman, 1994; Ruch, 1995). The responsible method factors have been held to comprise (a) ineffective or suboptimal *emotion induction* procedures (e.g. procedures that fail to induce the desired emotion in many participants, or that induce several mutually interfering emotions, rather than only the one intended); (b) suboptimal *measurement methods* of the syndrome components (e.g. insensitive measures; use of measures that refer to whole time spans rather than to a particular moment; use of strongly delayed verbal reports); and (c) inappropriate or suboptimal *data analysis methods* (e.g. inter- rather than intra-individual correlation analyses; neglect of possible nonlinear relations; failure to correct for measurement error, particularly in the self-reports). Evidence from several studies suggests that these method factors were indeed responsible for at least part of the low correlations among syndrome components obtained in many studies (see e.g. Lazarus, Speisman, & Mordkoff, 1963; Mordkoff, 1964; Opton, Rankin, & Lazarus, 1965; Rosenberg & Ekman, 1994; Ruch, 1995).

Objectives of the Present Research

The main aim of the research reported in this article was to explore the strength of the association among emotion syndrome components for *surprise*- held to be by many authors a biologically fundamental emotion (e.g. Ekman, 1972; Izard, 1977; Plutchik, 1980; Roseman, 1996)- while taking into account the described methodological concerns. Two additional, closely related aims were to examine the strength of the association between different subcomponents of the facial surprise expression (eyebrow raising, eye widening, mouth opening); and to test the hypothesis that the degree of coherence among the components and subcomponents of surprise increases with increasing emotion intensity (e.g. Davidson, 1992; Hodgson & Rachman, 1974).

The method of inducing surprise used in the present study, and the selection and measurement of surprise components, was based on a cognitive-psychoevolutionary model of the processes elicited by surprising events (see e.g. Meyer & Niepel, 1994; Meyer, Reisenzein, & Schützwohl, 1997b; Reisenzein, Meyer, & Schützwohl, 1996; Stiensmeier-Pelster, Martini, & Reisenzein, 1995). The core of this model is concerned with the *mental processes* elicited by surprising events. Briefly, it is assumed that a (ultimately) surprise-eliciting event initiates a series of processes that *begin* with the appraisal of this event as schema-discrepant or unexpected, *continue* with the occurrence of a surprise experience and, simultaneously, the interruption of ongoing information processing and the reallocation of processing resources to (i.e. the focusing of attention on) the schema-discrepant event, and *culminate* in an analysis and evaluation of this event plus- if deemed necessary- an updating or revision of the relevant schema. The first two steps in this series of processes are identified with the workings of the surprise mechanism proper, which is taken to be a phylogenetically old mechanism whose main evolutionary function is to monitor the person's schemas or belief system and help to update them in the face of unexpectedness (belief-disconfirmation, schema-discrepancy). Furthermore, it is assumed that most of the observable behaviours that have been traditionally associated with surprise, such as the interruption of ongoing motor action (e.g. Shand, 1920) or the facial display of surprise (e.g. Darwin,1872/1965; Ekman & Friesen, 1975) subserve, in one way or other, this evolutionary goal (Meyer, Niepel, & Schützwohl, 1994; Meyer et al., 1997b).

In accord with this model, surprise was induced in the present study by disconfirming some of the participants' beliefs. There are two principal ways to do this: first, one can disconfirm beliefs that were previously established in the experiment; second, one can disconfirm beliefs that the participants already hold (at least implicitly, in the sense of being disposed to form these beliefs when induced to reflect on the issue at stake). Previous studies of surprise have mostly relied on the first method, with the most frequently used induction method being a sudden, unannounced change of the mode of stimulus presentation after a series of no-change trials (e.g. Hiatt, Campos, & Emde, 1979; Landis, 1924; Meyer, Niepel, Rudolph, & Schützwohl, 1991; Parrott & Gleitman, 1989; Schützwohl, 1998; Siddle & Jordan, 1993). By contrast, the experimental paradigm used in the present study implements the second principal way of surprise induction. The participants worked on a computerised quiz task where they received immediate feedback about the correct solution. Some of the items had unexpected and hence surprising solutions (details are given in the method section). A major advantage of this method is that- because the quiz items can be chosen from widely different knowledge areas- surprise can be elicited repeatedly, which allows one to determine the association between syndrome components on an intraindividual basis. In contrast, the unexpected stimulus changes used to induce surprise in previous studies lose most of their surprise-eliciting capacity after their first or second appearance

(e.g. Reisenzein & Bördgen, 1998, exp. 4).
Four components of the surprise syndrome were measured. The *cognitive component* of surprise (the appraisal of unexpectedness) was measured prospectively, by asking the participants to select one of several possible answers to each quiz item and to state their confidence that the chosen response was correct. The *experiential component* of surprise (the feeling of surprise) was measured in the most straightforward way possible, by asking the participants to rate the degree of experienced surprise about the solution immediately after it was presented. The *behavioural (actionrelated)* component of surprise considered in this study consisted of surprise-related action delay. To measure this component, the presentation of the item solutions was embedded into a choice reaction time task: 700msec after the presentation of the solution, the participants had to decide as quickly as possible on the identity of two numbers.[2] Finally, I measured the *facial-expressive component* of surprise by recording the participants' facial reactions to the presentation of the item solutions and scoring them for the components of the surprise expression (e.g. Ekman & Friesen, 1975): eyebrow raising, eye widening, and mouth opening/jaw drop (see Method).

Previous Research

There are only few systematic studies of the associations among the components of the surprise syndrome, particularly studies in which more than two syndrome components were measured. The relation between the intensity of the *feeling* of *surprise* and the degree of *action delay* was examined in several previous studies of our research group (Meyer et al., 1991; Niepel, 1996; Niepel, Rudolph, Schützwohl, & Meyer, 1994; Schützwohl, 1998; Schützwohl & Reisenzein, 1998). The results were mixed. On the one hand, it was consistently found that the surprising event used in these studies- an unexpected change of the appearance of the distractors in a choice reaction time (RT) task- induced a reliable RT increase on the parallel task, as well as feelings of surprise. On the other hand, *within* the surprise groups, several studies found no reliable positive correlation between RT delay and the intensity of felt surprise, and occasionally even a slightly negative correlation was obtained (Meyer et al., 1991; Niepel et al., 1994). Other studies did find significant positive withingroup correlations, but they were always low to moderate only (.30- .40, Schützwohl, 1998; around .50, Niepel, 1996). Niepel (1996) also assessed *physiological reactions* (skin conductance responses, SCRs) to unexpected events in addition to feelings of surprise and response delay. In three studies, the physiological measures showed positive but low (.15- .39) and partly nonsignificant correlations to the other two variables. The SCR-RT correlations are in line with previous findings by Siddle and Jordan (1993).[3]
Finally, a number of studies examined *facial expressions* of surprise, mostly in children (Blurton Jones & Konner, 1971; Charlesworth, 1964; Ekman, 1972; Hiatt et al., 1979; Landis, 1924; Parrott & Gleitman, 1989; Wheldall and Mittler, 1976). These studies suggest low to moderate associations between the facial surprise expression and other syndrome components

---

2 Note that the action interruption/delay caused by unexpected events is not a ``pure' ' action, because at least its initial part is (presumably) not caused by an intention to stop ongoing processes but is an automatic, unavoidable effect of the appraisal of unexpectedness. However, for this reason the association of this surprise component to the other syndrome components should if anything be stronger than for a ``pure' ' action. Furthermore, action delay is not the only action-related component of the surprise syndrome, because the epistemic or investigatory processe s instigated by unexpected events (cf. the introduction) are (in my view) usually also goal-directed activities (cf. Meyer et al., 1997b; Stiensmeier-Pelster et al., 1995).

3 There is also a set of studies in which the correlations between different component s of the physiolog ical orienting reactionÐ which is usually held to occur both to novel and unexpected eventsÐ were examined (e.g. Barry, 1982, 1990; Barry & James, 1981). These studies, too, found at best weak associations among the different physiological variables.

(e.g. RT delay or *inferred* unexpectedness; a detailed review of these studies is contained in Reisenzein & Bördgen, 1998).

In sum, similar to the majority of studies of other emotion syndromes, the published studies on the degree of coherence of the surprise syndrome suggest at best weak associations among syndrome components. However, again like most research on other emotion syndromes, the existing studies on surprise were methodologically suboptimal for answering the question at issue. Some studies (e.g. Ekman, 1972; Landis, 1924) seem to have induced other emotions in addition to surprise, which could have obscured the component relations. The majority of the facial expression studies included no other surprise measures, but relied exclusively on the face validity of the unexpectedness manipulations. Yet other studies used suboptimal measurement points (e.g. ratings of experienced surprise were made only after a considerable delay; see Schützwohl, 1998). Finally, all studies computed the associations between syndrome components on a between-subjects rather than on an intra-individual basis. This may have introduced considerable noise due to theoretically irrelevant intraindividual differences (cf. Cacioppo et al., 1992; Ruch, 1995; see Results for an elaboration of this point).

Apart from potential methodological shortcomings of previous research on the surprise syndrome, there appears to be no published study in which the cognitive, subjective, behavioural, and expressive components of the surprise syndrome were assessed in a single experiment, allowing one to compute, for the same paradigm and the same participants, the complete matrix of intercorrelations among these components.

METHOD

Participants

The final sample consisted of 35 students (24 female) at the University of Bielefeld, with a mean age of 25.7 years (SD=5.2). Most were introductory psychology students who participated in partial fulfilment of their study requirements. Two additional participants were excluded from the data analyses, one because she denied permission to use her videotape, the other because she had 51% errors at the choice RT task.

Procedure

The participants were tested individually. They were told that the goal of the experiment was to investigate how people perform different but interlocked tasks. They would participate in a computerised quiz, in which a choice reaction time task- deciding as quickly as possible on the identity of two numbers- was embedded. Because I wanted to study surprise as much as possible in hedonically neutral form, care was taken that the participants would not interpret the quiz as an achievement situation: They were informed that the experimenter was primarily interested in the reaction time task, that the quiz was undiagnostic of intelligence or general knowledge, and that, to make the experiment more interesting and to provide them with an opportunity to learn something new, some of the questions had solutions that many people did not know. This instruction also served to provide the participants with a rationale for the surprise ratings (supposedly, these ratings served to determine more precisely how many people found the different solutions surprising), and to increase the likelihood that they would believe even highly unexpected solutions (which I take to be necessary for the elicitation of surprise). As a further means to achieve the latter goal, the participants were (truthfully) assured that, to the best of the experimenter's knowledge, all item solutions were correct.

The procedure was demonstrated to the participants with the aid of a practice item, after which they completed two additional practice items by themselves. After the experimenter had ascertained that the procedure was understood, the 45 experimental items were presented. To control for possible sequence effects, they were shown to each participant in a different, computer-generated random order.

Each quiz trial looked as follows. First, a quiz item (question) was presented on the computer monitor together with two to four response alternatives. For example, the participant was asked `Which of the following metals is the rarest?' ' , with the response alternatives being *platinum*, *uranium*, and *mercury* (Fig. 1A). The participant then selected what he/she believed to be the correct answer with the help of the mouse cursor (if all alternatives seemed equally likely, the participant was asked to guess). Subsequently, the question ``How certain are you about your answer?' ' appeared on the screen immediately below the response alternatives together with an 11-point rating scale with endpoints 0 = *not at all certain* and 10 = *completely certain*. Confidence ratings were entered by moving the mouse cursor to the appropriate scale number and clicking the button. Next, the screen was cleared, and the beginning of a sentence stating the correct solution was shown, such as ``The rarest metal (of platinum, uranium, mercury) is' ' : (see Fig. 1B). This ``solution phrase' ' was always complete with the exception of a single concluding word that denoted the correct answer. The solution sentence served both as a reminder (to avoid memory errors) and to introduce the presentation of the solution word. Participants were instructed to read the solution sentence and then to press the mouse button. This started the choice RT task, into which the presentation of the solution word was embedded. Simultaneously with the solution phrase, a frame had already been displayed at the centre of the screen (Fig. 1B). The participant's button press now caused a fixation cross, accompanied by a brief tone, to appear at the centre of this frame for the duration of 1400msec (Fig. 1C). This served

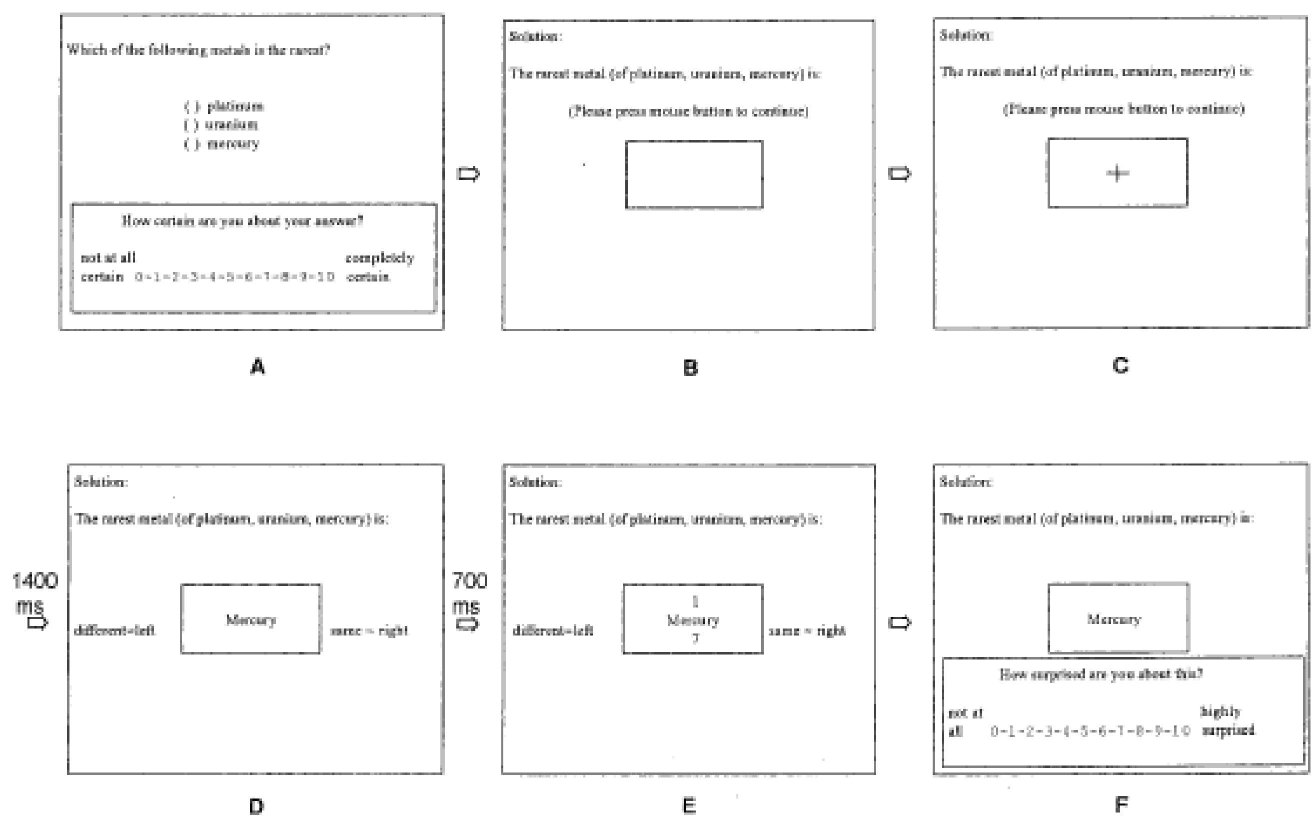

FIG. 1. Trial structure of the quiz paradigm.

to alert the participant that the solution word and the choice RT task were imminent. Immediately after the disappearance of the fixation cross, the correct answer was presented (e.g. *mercury* in the ``rarest metal' ' example; cf. Fig. 1D). Another 700msec later, two one-digit numbers were shown, one above and the other below the solution word (Fig. 1E). A somewhat longer SOA (700 rather than 500msec) than in previous studies using simple visual schema-discrepancies (e.g. Meyer et al., 1991) was used to allow for the semantic processing of the solution word (cf. Theios & Amrhein, 1989). Participants were asked to react as quickly as possible to the numbers by pressing the left mouse button when they were different and the right button when they were identical. To reduce memory load in the early trials, the stimulus-response rule (different-left, same-right) was also displayed on the screen. The numbers were selected randomly in each trial by the computer program with the restriction that they were different in half of the trials and identical in the other half. Both the solution word and the numbers remained on the screen until either the participant had reacted, or else for 7 seconds. Subsequently, a final question asking for the participant's feeling of surprise about the solution was presented, together with an 11-point rating scale ranging from 0 = *not at all surprised* to 10 = *highly surprised* (Fig. 1F). Responses were again entered with the mouse cursor.

To permit a fluent task execution, the participants did not have to confirm the answers before they were accepted by the computer. As a consequence, there was no possibility to correct erroneous entries. Participants were informed about this fact and asked to enter their responses with care.

Once the experimental trials had begun, the experimenter sat at a table located behind the participant and oriented 90 degrees away, where he/she busied him/herself with other tasks. Thus, the situation was minimally social in character, that is- apart from the fact that the experimenter rather than a stranger was present in the room- it fulfilled the criteria of a ``mere presence' ' paradigm (Guerin, 1986). This arrangement reflected a compromise between two conflicting intuitions about the effects of sociality on the expression of surprise. On the one hand, the presence of others could cause the suppression or masking of facial expressions (cf. Ekman, 1972). For this case, the level of sociality was expected to be low enough to minimise this inhibitory tendency, as supported by other research (Ruch, 1995). On the other hand, it seemed conceivable that, if the surprise expression serves mainly communicative purposes, it occurs preferably or only in a social situation (Reisenzein et al., 1996). For this eventuality, the level of sociality was hoped to be high enough to allow a surprise display to occur, as again suggested by other research (e.g. Chovil, 1991; I return to this point in the Discussion).

At the end of the study, the participants were informed about the true purpose of the experiment and the fact that they had been videotaped and were asked for their permission to analyse the recordings. As mentioned, only one participant declined.

Expression Recording and Coding

The computer monitor and the mouse were located on a table positioned toward the middle wall of the experimental room in front of a doublesheeted glass window. The adjoining soundproofed room contained a video camera fitted with a zoom lens, plus recording equipment. The camera was mounted on a tripod in such a way that it took a frontal picture of the participant's head and shoulders slightly from above. The camera's path of vision up to the window was surrounded with heavy black cloth, and the camera room was darkened, whereas the experimental room was illuminated by ceiling lights and a table lamp, both of which reflected in the window pane. This arrangement completely obscured vision of the contents of the camera room. To diffuse possible remaining suspicions of the participants, the computer monitor and the table on which it was placed were oriented in a flat angle towards the wall (hence the participant looked slightly by the side of the window), and the window pane was covered with posters and paper sheets that left only a small gap in the camera's path of vision. To synchronise the videorecording to the experimental events, the participant's monitor was connected via a switch to a second monitor located in the recording room, where the central part of the screen image was filmed by another camera (only the central portion was recorded to enhance visibility). This video image was inserted with the aid of an electronic mixer into the main recording (showing the participant's heads and shoulders), where it appeared as a small insert at the right bottom corner of the screen. Finally, the videorecorder was connected to a microphone hidden in the experimental room.

The videotapes were coded by a student with the help of a coding scheme developed in the course of a different study using the quiz paradigm (Mondkowsky, 1996). The rater indicated for each quiz item whether or not the expression categories were manifested in the interval beginning with the onset of the fixation cross (which immediately preceded the presentation of the solution word) and ending with the offset of the surprise rating scale (which marked the start of the next trial). To ensure blindness of the rater to the identity of the items to which a participant responded, the insert showing the events on the participant's monitor was covered with a sheet of paper that left only a small gap through which the bluish frame of the surprise scale, but not the scale points nor the check marks made by the participant, could be seen. Disappearance of this frame marked the end of the coding interval for an item, whereas its beginning was signalled by the tone accompanying the fixation cross.

Coding System. A present/absent coding scheme with eight categories was used. Four categories referred to facial or vocal surprise displays. A ``full-fledged' ' facial expression of surprise was defined to consist of three components (see Darwin, 1872/1965; Ekman & Friesen, 1978; Frois-Wittmann, 1930; Wallbott & Ricci-Bitti, 1993): raising of the eyebrows (action units AU1/AU2 of Ekman and Friesen's, 1978 Facial Action Coding System), widening of the eyes (accomplished by raising the upper eyelid, AU5), and opening of the mouth/jaw drop (AUs 26/27). Each of these three facial expression components was coded as present if it was visible in the slightest degree within the coding time window. A fourth coding category captured surprise vocalisations, typically consisting of the uttering of ``oh' ' or ``wow' ' . Mouth opening was coded in conjunction with a surprise vocalisation only if it clearly preceded, and thus was apparently not simply the by-product of, the vocalisation (however, vocalisations occurred only 14 times in the total 1575 cases). The remaining three coding categories comprised smiling/laughter (AU12) as well as various nonverbal and verbal responses reflecting cognitive responses to the solution words: nodding; verbal acknowledgment of the solution (``aha' ' ); and affirmative vocalisations (e.g. ``see' ' ; the verbal response categories were coded as mutually exclusive because they virtually never co-occurred). These additional categories were suggested by a preliminary examination of the tapes from this and a different study using the quiz paradigm (Mondkowsky, 1996).

Two kinds of data were available to estimate the reliability of the coding system. First, in the course of time measurements, the original codings of five high-expressive participants of the present study (making for a total of 225 trials) were checked by a second coder. The agreement between the two coders obtained in this way on any one coding category was ≥ 97% (≥ 88% when corrected for chance, Cohen's, 1960, Kappa χ 100). Second, the person who coded the tapes from the present study also coded the videos from a different study using the quiz paradigm (Mondkowsky, 1996); the tapes of eight participants from this latter study (totalling 238 quiz trials) were later scored by another coder who was blind to the original codings. Agreement between these codings was 98% (93% chance-corrected) for brow raising, 99% (83%) for eye widening, 96% (83%) for surprise vocalisations, 98% (81%) for nodding, 95% (73%) for smiling/laughter, and 97% (54%) for mouth opening/jaw drop. The reliability estimates for the last category must however be treated with caution, because they were based on only 8 or 10 (depending on coder) positive cases; the first method of determining intercoder agreement suggests that this estimate is too conservative. The reliability of the combined index of facial expression used in the main analyses reported below (consisting of the sum of the three surprise expression components) was *r*= .98 and .89 for the two methods of reliability assessment.

Quiz Items

A total of 45 quiz items were presented. They were compiled from almanacs, the game *Trivial Pursuit*, encyclopedias, statistical yearbooks, and similar sources.[4] Ten items, including most of the highly surprising ones, were selected on the basis of surprise ratings collected in a previous questionnaire study involving 60 participants. The remaining items were selected from a newly constructed item pool on the basis of their classification, by 10 students, as ``very surprising' ' versus ``not or only mildly surprising.' ' My aim was to include only about 20% highly surprising items to avoid the build-up, during the course of the experiment, of the belief that most of the questions had unexpected answers. Once established, this belief could have changed the participants' reactions to the items yet to come. That is, once this belief had been created, the participants could have used the seeming improbability of an answer as a cue to its correctness. Although this need not have eliminated surprise, it could at least have shifted its occurrence from the point of solution presentation to that of item presentation.

Dependent Variables and Data Treatment

For each quiz item, the computer recorded the participant's answer, his/her judged confidence in the answer, response latencies on the choice RT task, and the rated intensity of surprise about the solution.

The first two variables were subsequently combined into a *prospective* estimate of the expectedness versus unexpectedness of the solution as follows: if the participant had chosen a wrong answer (e.g. *platinum* in the ``rarest metal' ' example), the degree of unexpectedness of the subsequently presented correct answer (here, *mercury*) was set equal to the confidence rating; if the correct answer had been chosen, the confidence rating was multiplied by - 1. This resulted in a scale ranging from - 10 to +10. The lower end represents certainty about the correct answer and hence maximum expectedness of the solution presented later; 0 represents maximum uncertainty; and the upper end represents certainty about a wrong answer and hence high unexpectedness of the solution.[5]

As concerns the RT data, responses that deviated more than three standard deviations from a participant's mean were replaced by the respective mean to reduce the effect of outliers (e.g. Fazio, 1990). This resulted in the replacement of maximally three responses per participant (M = .9). Alternative outlier treatments (elimination, fixing at 3z) produced virtually identical findings.

The three components of the facial surprise expression (brow raise, eye widening, mouth opening/jaw drop) were summed into an overall index of surprise ranging from 0 (no component shown) to 3 (all three components shown). Preliminary inspection of the tapes revealed that 17 of the 35 participants put their fingers or hand to their chin or mouth, or rested

---

4 A list of the quiz items, including the item difficulties and the means and standard deviations of the four dependent variables, is available from the author.

5 In the case of a wrong answer, this estimate of unexpectedness is completely adequat e only for dichotomous items and for multiple choic e items with subject ively equiprobable unchosen answers. For other multiple choice items, the method leads to an underes timation of unexpectedness in the case of a wrong answer. I decided to buy this imprecision of the unexpectedness measurement, because the feasible alternatives (e.g. using ony dichotomous items, requesting certainty ratings for all answer alternatives, or asking the participants to type in their own solutions) seemed even less attractive. However, the high intra-individual correlation between unexpectedness and self-rated surprise suggest s that this source of error was small.

their chin on their hand during at least part of the experiment, which made the detection and possibly even the occurrence of mouth opening difficult. Indeed, these participants had fewer ``mouth opening' ' codes (2.6%) than the remaining 18 participants (4.0%), although this difference is not statistically significant [$t(33) < 1$]. However, the three-component expression index correlated nearly perfectly (.96 in the pooled data) with a twocomponent index comprising only eyebrow raising and eye widening, whose relations to the other variables were also virtually identical to those obtained for the three-component index. Therefore, with the exception of the analyses of the association between the individual expression components, I report the results for the three-component index based on the total sample. Note that this sum index is ``liberal' ' in that it (a) counts any single component of the facial surprise expression, as well as any combination of components, as an expression of surprise; (b) allows that surprise is expressed by different (combinations of) components in response to different items or by different participants; and (c) is sensitive to the possibility that increased surprise intensity is reflected in a greater number of components (for more on these issues, see e.g. Darwin, 1872/1965; Ekman & Friesen, 1975; Frois-Wittman, 1930; Wallbott & Ricci-Bitti, 1993). Vocal surprise expressions (which, as said, occurred only 14 times) were not included into the surprise index because my main interest was on facial expression and because I was uncertain about the relation of vocal expressions to the other expression components. In fact, they turned out to be nearly uncorrelated with the facial expression components (see Results).

The major data analyses thus involved the following four measures of surprise: (a) the estimated unexpectedness of the quiz answers, calculated as described above (a fairly straightforward indicator of the cognitive component of the surprise syndrome); (b) ratings of the intensity of surprise felt in response to the presentation of the item solutions (the most direct possible indicator of the feeling component of surprise); (c) response time at the parallel choice RT task (a direct measure of the behavioural component of surprise, understood as observable action delay); and finally (d) a composite index of facial surprise expressions (a measure of the facial-expressive component of the surprise syndrome).

I believe that the described experimental paradigm, together with associated data analysis methods (see Results) effectively controls for most nuisance factors that could have reduced the association between the syndrome components of surprise in previous studies (cf. the Introduction).

1. The paradigm allows the induction of surprise of fairly ``pure' ' *quality*, that is, uncontaminated by other, simultaneously present emotions. The only other emotion that seems to have been elicited with some frequency by the quiz items- as suggested by the occurrence of smiling or laughter in response to some items- was amusement (see the Results). Furthermore, the paradigm allows the induction of different surprise *intensities* (see Results), as well as precise control over the *timing* of the surprise induction.

2. Problems associated with the reference and timing of the *measurement* of the surprise syndrome components are also largely avoided. Each measurement refers to a temporally and contextually clearly specified, potentially surprising event (the presentation of a particular quiz item solution). Furthermore, the self-reports are collected either immediately before (assessment of belief strength) or immediately after (surprise rating) each surprising event, minimising possible memory errors.

3. In contrast to previous studies of the surprise syndrome, data from a sufficient number of surprise incidents are available for each participant to permit intra-individual analyses of syndrome coherence, with all attendant benefits (see Results).

Still other potential problems of previous studies- in particular the failure to take into account possible nonlinear relations and to correct for measurement error in self-reports- were addressed by means of special data analysis procedures (see Results).

RESULTS

The data were analysed as follows. First, as a preamble to the main analyses, I verified that the different quiz items were effective in inducing a sufficient range of degrees of surprise, and tested for possible trial effects. Second, I analysed the strength of association between the four measured components of the surprise syndrome, using an intra-individual analysis strategy. Third, additional analyses addressed the issues of possible nonlinear relations and of measurement error in the self-reports. Fourth, I tested the hypothesis that the degree of coherence among the surprise components increases with increasing emotion intensity. Fifth, I analysed the association between the components of the facial expression of surprise, and retested the intensity-coherence hypothesis in this more restricted context.

Effectiveness of the Paradigm

Before turning to the association among the surprise measures, it is important to verify that the quiz items were effective in inducing different degrees of surprise in the participants. This question was examined for both items and participants as units of analysis.

Between-subjects (within-items) analyses. Summary information on the means and SDs of the four surprise measures for the different quiz items is contained in the left half of Table 1. As can be seen, the item solutions differed greatly in average unexpectedness and mean rated surprise, ranging from the fully expected (M= -9.8) and unsurprising (M = .1) to the fairly unexpected (+4.9) and surprising (7.1). The items also differed considerably with respect to mean RT on the parallel choice RT task (ranging from about 700msec to about 1000msec) and facial expression (from 0 to .57); however, even the most strongly ``expression-drawing' ' item caused a facial surprise expression (at least one component) in only 34% of the participants.

For a formal test of the statistical significance of the between-item variation, the 45 items were divided into three groups of 15 items comprising the comparatively least, medium, and most highly unexpected items, and the means of the four surprise measures in the three groups were compared by repeated-measures ANOVAs. To guard against violations of sphericity, the *df*s of the *F*-tests were adjusted using the Huynh -Feldt ε (see Vasey & Thayer, 1987). All comparisons were highly significant, all [$Fs(2,68) > 12$, $Ps < .001$, minimum Huynh- Feldt $\varepsilon = .61$].[6] Moving from

[6] I report the uncorre cted *df*s and corresponding *P*s, plus e , but all repeated-measures ANOVA tests declared significant remain so after correction.

TABLE 1
Effectiveness of the Paradigm (for Items and Participants)

	Items (N = 45)				*Participants (N = 35)*			
	min	*max*	*Median*	*Mean*	*min*	*max*	*Median*	*Mean*
Unexpectedness (−10 . . . +10)								
Mean	−9.8	4.9	−.9	−1.7	−3.5	−.44	−1.6	−1.7
SD	.7	7.4	4.4	4.3	4.7	7.8	6.4	6.4
Surprise Rating (1 . . . 10)								
Mean	.1	7.1	3.8	3.5	.89	5.9	3.6	3.5
SD	.5	3.7	2.8	2.7	2.0	4.0	3.2	3.2
Items rated as highly surprising (≥ 7)	–	–	–	–	3	25	10	10
Response Time (msec)								
Mean	698	1073	903	900	596	1622	846	900
SD	216	626	323	341	93	874	229	254
Surprise Expression (0 . . . 3)								
Mean	.00	.57	.14	.19	.00	.96	.11	.19
SD	.00	.95	.53	.49	.00	1.3	.36	.40
% Participants/Items with Expression[a]	0	34	11	12	0	60	9	12

[a] *For items*: Percentage of participants who showed at least one surprise expression component. *For participants*: Percentage of items to which at least one surprise expression component was shown.

low to medium to highly unexpected items, there was a near perfect linear increase in average unexpectedness (- 7.5; - .97; 3.2), felt surprise (1.1, 4.0, 5.3), response time on the parallel task (861msec, 918msec, 966msec), and facial expressiveness (.05, .19, .31).

Within-subjects (between-items) analyses. As shown in the right half of Table 1, there were also substantial interindividual differences in the reactions to the quiz items; for example, the mean surprise ratings (across all items) of a participant ranged from .9 to 5.9. However, as judged by the surprise ratings, all participants were fairly strongly surprised by at least some items (as well as unsurprised by others): ratings ≥ 7 were given by a participant to, on average, 10.7 (24%) of the items (SD = 5.3, min = 3, max = 25); and ratings ≥ 8 were given, on average, to 7.8 (17.3%) of the items (SD = 5.1,min = 0, max = 23). The attempt to expose the participants to about 20% highly surprising items was therefore by and large successful. With the exception of the facial expression index, all participants showed sufficient variance on all surprise measures to be included in the subsequent, intra-individual correlational analyses. Eight of the 35 participants failed to show any facial surprise expression component, and five more showed an expression to fewer than three items. To maximise the comparability of the findings to those of the methodologically similar study by Ruch (1995) (concerned with the relation between smiling and amusement), these 13 participants (37% of the sample) were excluded from the later within-subjects correlational analyses involving facial expression. However, it should be noted that there was no evidence from the three remaining surprise measures that these participants were less surprised than the rest, all *ts*(33) < 1. I come back to this point in the Discussion.

Trial Effects. As mentioned in the Method section, there was a concern that the participants might, in the course of the experiment, develop the belief that most of the items had unexpected answers, which could have attenuated their surprise reactions. Although such an effect, unless it is so strong that it results in a substantial reduction of variance, would not affect the *correlations* between the surprise measures, it was of interest to check whether this effect occurred at all and if yes, how strong it was. For this purpose, the 45 trials were grouped into three blocks (1- 15, 16- 30, 31- 45), and the means of the four surprise components in the three blocks were compared by repeated-measures ANOVAs. No significant block effects were obtained for felt surprise and facial expression; significant effects were, however, obtained for unexpectedness [$F(2,68) = 4.4$, $P < .05$, Huynh- Feldt $\varepsilon = .96$], and for RT [$F(2,68) = 13.3$, $P < .001$, Huynh-Feldt $\varepsilon = .62$]. Follow-up *t*-tests showed that: (a) the solutions were on

average significantly more unexpected in the first block (2 1.1) than in the second (2 2.1) and the third block (- 2.0) [*t*s(34) ≥ 2.6, *Ps* < .05], whereas the second and third block did not differ; (b) mean RT decreased significantly from the first block (1002msec) to the second block (864msec), [*t*(34) = 3.6, *P* < .01], and again marginally significantly from the second to the third block (834msec), *P* <.08.
However, because the most rapid RT decline occurred during the first five trials, the RT decrease across blocks reflects most likely a practice effect rather than a decrease of unexpectedness (e.g. during the first few trials the participants may have repeatedly verified the response key assignment). Supporting this interpretation, an analysis of covariance of RT, with unexpectedness statistically held constant, replicated the original block effect, [*F*(2,67) = 13.1< *P* , .001].
In sum, three of the surprise measures gave no (surprise rating, facial expression) or no unambiguous (RT) support for a reduction of the surprise potential of the items across the experiment, and the block effect obtained for the fourth measure (unexpectedness) was very small (one scale point on the 21-point unexpectedness scale). In addition, computations of the correlations among the four surprise components for the first, second, and third block (conducted for the data pooled across items and participants) provided no evidence for a decrease of correlations across blocks.
On the contrary, four of the six correlations *increased* slightly (on average by .08) from the first to the third block, and the remaining two increased slightly from the first to the second block.

Strength of Association between the Surprise Components

Ruch (1995; see also Ruch, 1990) argued- and provided empirical support for this claim- that the inconsistent and often low correlations obtained in previous humour studies between self-report indicators of exhilaration on the one hand, and facial mirth expressions (smiling/laughter) on the other hand, were partly artefacts caused by the method used to compute the correlations (for an earlier, analogous argument in the area of stress research, see Lazarus et al., 1963). That is, in different studies, feelingexpression correlations were computed either between or within participants (i.e. nomothetically vs. idiographically), and either for raw data or for data that were aggregated across items or participants; furthermore, suboptimal data analysis designs- in particular a between-subjects rather than a within-subjects design- were apparently often used. I took this methodological consideration into account by focusing on *intra-individual* correlations among syndrome components for *nonaggregated (raw) data* as the method of choice. That is, for each pair of surprise measures X and Y, the correlation between X and Y was computed separately for each participant across the 45 items, and the average correlation was determined using the Fisher *r* to *z* transformation (in the case of the facial expression index, only the 22 participants who showed a facial surprise reaction to at least three items were included).
Given the present data set (35 subjects responded to 45 items) several other analysis methods are possible (see Ruch, 1995; Reisenzein, 1999): computing the average within-item (between-subjects) correlation; computing the correlation for the pooled (35 X 45 = 1575) raw data; correlating the subject means (the mean X and Y responses of each participant across the 45 items); and correlating the item means (the mean X and Y reactions of the 35 participants to each item). Two of these alternative designs, the betweensubjects and pooled raw data design, were used in the context of testing the intensity-coherence hypotheses, and to determine the association between the components of facial expression. In both cases, this was done because I felt that there were not sufficient data to test these hypotheses at the intraindividual level. However, for answering the central question at stake- the covariation of the surprise syndrome components for a person at a particular moment in time- the alternative designs are either suboptimal (the betweensubjects and pooled design) or even unsuited (the aggregate designs).
Briefly, the reasons for this are as follows (for more detail, see Cacioppo

et al., 1992; Reisenzein, 1999; Ruch, 1995). (a) Aggregation across subjects (i.e. the correlation between the item means) focuses on the wrong level of analysis, the group rather than the individual. The remaining analysis methods do have the individual as the unit of analysis. However, (b) aggregation across items (i.e. correlation of the subject means) is not concerned with the association of the components of a person's emotional reactions at a given point in time, but with the association between a person's response tendencies or dispositions. Although of interest in its own right, this question is peripheral to the present concerns. (c) The between-subjects (within-items) raw data design does address the association between the components of surprise episodes, but it most likely underestimates the strength of this association, because it fails to eliminate inter-individual differences in surprise measures that are irrelevant from a theoretical point of view. Apart from response styles (in the case of the self-reports) and differences in baseline level of responding (in the case of RT), these differences concern in particular theoretically permissible variations in the form and the parameters of the functions that relate the surprise measures (e.g. the slope and intercept in the linear case; cf. e.g. Cacioppo et al., 1992). In line with this reasoning, I obtained, replicating Ruch (1995), consistently higher positive and fewer negative correlations with the intra-individual than the inter-individual analysis method (see Reisenzein, 1999, for details). (d) Finally, the pooled raw data design combines the intra- and inter-individual analysis methods; therefore, this design suffers partly from the same problems as the intra-individual design.

TABLE 2
Association among the Syndrome Components

	Correlations[a]					
	UE-Sur	*UE-RT*	*UE-Exp*	*Sur-RT*	*Sur-Exp*	*RT-Exp*
N coefficients	35	35	22	35	22	22
Pearson r—mean[b]	.78	.35	.41	.32	.46	.19
Pearson r—median	.78	.33	.36	.30	.38	.23
Nonlinear Relations						
3rd degree polynomial fit (R)[c]	.82(.80)	.43(.35)	.45(.39)	.41(.33)	.45(.38)	.30(.20)
Spearman Rho	.80	.36	.36	.35	.36	.24
Kruskal Tau	.59	.24	.14	.21	.14	.09
Measurement Error in Self-reports						
Partial aggregation of self-report data	.91	.33	.37	.29	.36	.23
Correlations for hypothetical error-free self-reports	1.00	.49	.48	.46	.49	.23

[a] UE, prospective estimate of unexpectedness: Sur, surprise rating; RT, response time on parallel task; Exp, facial expression of surprise (number of surprise components shown, 0 to 3). The table shows the mean (line 2) or median (remaining lines) of 35 or 22 within-subject association coefficients (only 22 of the 35 participants showed a facial surprise expression to at least 3 items).
[b] Assuming independent data points (N = 45 items), individual rs $\geq$.30 and $\geq$.39 would be significant at α = .05 and .01, respectively.
[c] Coefficients in parentheses are adjusted multiple Rs.

These considerations leave the within-subjects raw data analysis design as the best-suited design for answering the research question at issue. Like the between-subjects analysis for raw data, it is targeted at the correct unit of analysis, the momentary emotional episodes of individuals. In addition, it eliminates error variance due to theoretically irrelevant level and scatter differences between individuals in X and Y responding, and it allows for theoretically permissible inter-individual differences in the functions relating the surprise measures (meaning, for linear functions, individual differences in slope and intercept).

Line 2 of Table 2 shows the average intra-individual Pearson correlations among the four surprise measures. Discussion of the results will focus on the size of the correlation coefficients rather than on their statistical significance, because the former is of main interest; however, I also report the significance tests for the *mean* intra-individual correlations.[7] The *t*-tests of the mean (*z*-transformed) correlations (with *df*s of 34 and 21, respectively) showed that all mean *r*s were signficiantly ($P < .001$) greater than zero. (The distributions of the coefficients across participants were in four cases about normal, and in the remaining two cases- unexpectednesssurprise and expression-RT- slightly left-skewed). However, only one of the six coefficients- that describing the association between the two mental syndrome components, prospectively estimated unexpectedness and retrospectively judged surprise- was high (mean $r = .78$). The remaining correlations were moderate to low (.46- .19). The highest of these correlations was obtained for the association between the two mental syndrome components and facial expression (.41, .46). This was followed by the correlation of the mental components to RT (.35, .32) and the correlation between facial expression and RT (.19). Pairwise multiple comparisons of the coefficients using dependent *t*-tests (with *df*s of 34 or 21, respectively) and α adjusted for the number of tests (.05/15) showed that the unexpectedness-surprise correlation was significantly higher, and the expression-RT correlation significantly lower, than each of the other coefficients.

Role of Nonlinearity and Measurement Error in Self-Reports

The just-reported analyses still suffer from two methodological drawbacks mentioned in the Introduction. First, they do not take into account that- in contrast to the behavioural and expressive components of the surprise syndrome (understood as observable responses)- the *mental* components of the surprise syndrome cannot be directly measured, but can only be indirectly assessed via self-report instruments that are subject to observational and reporting errors (see the Introduction, and Rosenberg & Ekman, 1994). These (nonsystematic[8]) measurement errors lower the associations of the mental components of the surprise syndrome to each other and to the behavioural and expressive components. Second, whereas the preceding analyses presupposed that the syndrome components are related linearly, emotional syndrome theory allows for nonlinear (positive) monotonic relations. Although linear regression usually provides for a good approximation to nonlinear monotonic functions (e.g. Dawes, 1979), particularly with errorful data, it underestimates the strength of association between variables in cases of pronounced nonlinearity. These two issues- nonlinear relations and measurement error in the self-reports- were addressed in the following subsidiary analyses.

Nonlinearity. To test for the existence of nonlinear monotonic relations among the syndrome components, the previous analyses were repeated using three alternative indices of association that are, each in its own way, sensitive to such relations: (a) the multiple *R* of a third-degree polynomial regression (Neter Wasserman, 1974), which allows to approximate positively and negatively accelerating as well as sigmoidal (acceleratingdecelerating) functions (see e.g. Cacioppo et al., 1992). In these analyses, unexpectedness served as the independent variable for all other variables, the surprise rating was the independent variable for RT and facial expression, and RT was the independent variable for expression.[9] (b) Spearman

---

7 There appears to be no generally applicable statistical test for *single* intra-individual correlations (where the data points cannot be taken to be statistically independent). As a rough guideline, I report the individual correlations assuming independence (Tables 2 and 3).

8 Systematic measurement error can also lead to an *increase* of the correlations. As mentioned, some sources of systematic error (e.g. verbal report styles such as acquiescence, cf. Green, Goldman, & Salovey, 1993) were eliminated by the use of an intra-individual design.

9 This choice was motivated by the desire to use the more nearly continuous variables (unexpectedness, surprise, and RT) as predictors, because this allows to obtain a better fit. Otherwise the choice of predictors was arbitrary.

Rho and (c) Kendall Tau rank correlations, both of which allow the fitting of arbitrary monotone functions (Rho in contrast to Tau assumes equal intervals between ranks).

The results of these analyses are shown in lines 4 to 6 of Table 2. In contrast to the previous analysis, the medians of the intra-individual association coefficients are reported, because the Fisher *z*-transformation cannot be applied to the polynomial *R* and Tau. To allow a direct comparison with the earlier linear analyses, the median Pearson correlations are also listed (line 3).

Pairwise nonparametric (rank) comparisons of the six Pearson correlations with the corresponding polynomial *R*, Rho, and Tau coefficients, using the sign test (Bortz, Lienert, & Boehnke, 1990) and a adjusted for the number of tests (.05/18), showed that the Polynomial *R* was significantly higher in all cases; Tau was significantly lower in all but one case (RT-expression), and Rho was significantly higher in one case (unexpectedness-surprise). The maximum increase obtained for the association coefficients was + .11 for the polynomial *R* and + .05 for Rho and concerned in both cases the surprise-RT association. The results for the polynomial regression must however be regarded with caution, because- given that these coefficients are based on a fitted function with four estimated parameters- the higher *R*s may in part only reflect capitalisation on chance fluctuations in the data. The adjusted coefficients, shown in parentheses, may be more realistic estimates (maximum improvement: 1 .03). It may be concluded from these analyses that some nonlinearity does exist, but is either so small or so much blurred by noise that the nonlinear association coefficients perform only little better than the linear correlations.

Measurement Error in the Self-report Data. To eliminate at least part of the random measurement error in the self-report data, I used a partial across-item aggregation procedure based on the idea of pseudo-replication (or parallel tests). For each participant separately, the 45 items were first rank-ordered by unexpectedness from 1 to 45 (ties were broken randomly), and three-point moving averages (i.e. the means of items 1- 3, 2- 4, and so on up to 43- 45) were computed for unexpectedness and rated surprise. The intra-individual correlations among the syndrome components were then recalculated for these aggregated scores. This procedure is based on the (certainly not entirely unproblematic) assumption that different quiz items that are about equally unexpected for a given person can be treated as approximate replications of one and the same item for this person. Therefore, by aggregating the subjective measures across the replication items, measurement error is reduced, whereas substantive variance is retained.

The median intra-individual coefficients obtained by this procedure are shown in line 7 of Table 2. As can be seen, the only noteworthy increase of coefficients, relative to those for the nonaggregated data shown in line 3, occurred for the correlation between unexpectedness and surprise, which rose from .78 to .91 (only this difference was significant by the sign test with $\alpha = .05/6$). Although this attests to the effectiveness of the aggregation procedure for the reduction of measurement error, this reduction of error had next to no effect on the strength of association between the mental and behavioural syndrome components. These results suggest that errorful selfreports were *not* an important reason for the low correlations among the mental and behavioural syndrome components.

This conclusion receives additional support from the following theoretical consideration: Given the original correlation between unexpectedness and rated surprise (on average, .78), the lower bound of the reliability of these measures- that is, the reliability which they would have if the underlying mental states were perfectly correlated (which, to note, is not necessarily predicted by the theory of surprise processes described in the Introduction) and assuming that errors are uncorrelated- is of the same size. Consequently, the maximum increased in the median correlations between mental and behavioural syndrome components that could theoretically be obtained by correcting for unreliability in the subjective measures would be .16 (for surprise-RT; see line 8 of Table 3), and the highest attainable correlation would be .49 (between the two mental components and expression).[10]

Coherence among Surprise Components as a Function of Intensity

Several authors have proposed that the degree of coherence among components of emotion syndromes might increase with increasing emotion intensity (e.g. Davidson, 1992; Hodgson & Rachman, 1974). This hypothesis is a crucial auxiliary assumption in attempts to attribute low association coefficients to insufficient surprise intensity (cf. the Discussion). Therefore, and because the hypothesis is of interest in its own right, it was tested in additional analyses. Because I thought that there were too few data for testing the hypothesis at the intra-individual level, and because primarily the *relative* size of the association coefficients is of interest in this case, the tests were made using the between-subjects and pooled draw-data designs.

In the first test, the mean unexpectedness scores of the 37 quiz items with sufficient variance (or, in the case of facial expression, the 27 sufficient-variance items to which at least three participants reacted) were correlated with the Fisher *z* transforms of the six correlations between syndrome components obtained from the different items. In conflict with the intensity hypothesis, the correlations between degree of unexpectedness and strength of syndrome coherence were, with one exception (the facial expression-RT association, $r= .09$) *negative*, ranging from - .18 (the surprise-RT association) to - .46 (the unexpectedness-RT association, $P < .05$). Parallel findings were obtained when rated surprise was used as the index of emotion intensity. Hence, these analyses suggest, if anything, a tendency toward *less* syndrome coherence at higher levels of emotion intensity. This finding cannot be attributed to a reduction of variance at higher intensity levels: although the SD of unexpectedness showed a slight tendency to decrease with increasing intensity ($r = -.33$, $P= .09$), this was not the case for the other three variables, the surprise ratings ($r = .02$), RT ($r = .57$, $P < .01$), and facial expression ($r = .88$, $P <$ .001).

In the second test, the pooled raw data where split into four groups according to rated surprise intensity (0- 1; 2- 4; 5- 7; 8- 10) and the associations among syndrome components were computed separately for the four subgroups. The average correlations among the syndrome components in the four groups were .18, .18, .09, and .11.

Analysis of Expression Components

The data collected in this study also afforded an opportunity to analyse the association between different components of the facial surprise expression, as well as the relation of these components to the other expression codes. To avoid bias, these analyses are reported for both the whole sample and for those participants ($N= 18$) who did *not* put their fingers or hand to their chin or mouth, or rested their chin on their hand during the experiment

---

10 With respect to the associat ion of facial expression to the other variables, it could also be argued that the linear correlation was deflated by the fact that the expression index was strongly skewed for most participants (cf. Ruch, 1995). Apparently supporting this argument, I found a significant association between the frequency of surprise expressions shown by a person and the correlat ion of expression with self-rated surprise [*r*(*N* 5 22) 5 .42, *P* , .05]. Nevertheless, I propose to treat this argumentÐ and attendant proposa ls for an upward correction of the linear coefficientsÐ with caution. My main reason is as follows: The deeper reason for the restrict ion of the linear correlation by skewness (and other shape) differences are nonlinear relations between the variables (e.g. Carroll, 1961; Karabinus, 1975). This possibil ity was already covered by the preceding tests for nonlinearity; the results of these tests indicate against a skewness correction of the linear coefficients.

(cf. method). The most frequently observed surprise expression component was eyebrow raising, shown in 150, or 9.5% of the total 1575 cases (53, or 6.5% of the 810 cases of the subsample). This was followed by eye widening (5.8/3.7%), mouth opening (3.3/4.0%), and surprise vocalisation (.9/.5%). Of the remaining coding categories, smiling/laughter occurred in 4.4/4.3%, nodding in 3.0/2.1%, and affirmative vocalisations (``see!' ') in .2/0%.

Association among Expression Components. Because the individual expression categories were thought to occur too infrequently to allow meaningful within-subjects analyses, the correlations for the pooled raw data (i.e. pooled across participants and items) were computed. Six categories were observed with sufficient frequency to allow inclusion into these analysis: the facial and vocal surprise categories, smiling/laughter, and nodding. The correlations among these categories are shown in Table 3.

TABLE 3
Associations among the Expression Categories[a]

Categories	*1*	*2*	*3*	*4*	*5*
1. Brow Raise					
2. Eye Widening	.63(.64)				
3. Mouth Opening	.32(.38)	.41(.50)			
4. Surprise Vocalisations	.13(.13)	.06(−.01)	.14(.17)		
5. Smiling/Laughter	.11(.07)	.05(.02)	.08(.02)	.18(.07)	
6. Nodding	.08(.10)	.10(.11)	.07(.15)	−.02(−.01)	.02(.01)

[a]Phi coefficients for the pooled data, based on either the total sample (N = 1575) or (in parentheses) a subsample of 18 participants (N = 810; see text for an explanation). Assuming independent data points, $rs \geq .05$ and $\geq .07$ would be significant at α = .05 and .01, respectively, for N = 1575.

The results suggest moderate correlations among the three facial surprise expression components (largest r = .64) and weak to absent correlations to and within other expression categories. As these coefficients show, the components of the facial surprise expression frequently did not occur simultaneously. On the contrary: In the majority (54/56%) of the 182/70 cases in which a surprise expression was observed, only one of the three components of the full-fledged expression was coded. Two components were coded in 31/24% of the cases (most were co-occurrences of brow raising and eye widening), and the full three-component surprise face was shown in 15/20%. Finally, 9 of the 14 surprise vocalisations in the total data set occurred together with one or more facial expressions, whereas 5 occurred alone.

Coherence among Facial Expression Components as a Function of Intensity. Analogous to the earlier analyses for the relations between different syndrome components, I examined whether the coherence of facial expression components, and hence the probability of a ``full-fledged' ' surprise expression, increases with surprise intensity. To test this hypothesis, the items were again grouped into four categories according to rated surprise intensity, and the frequency with which one, two, or all three components of the facial surprise expression were shown was calculated for each group (parallel analyses for estimated unexpectedness gave virtually identical results). For the lowest surprise intensity group (0- 1), these percentages were (2.3, 0.5, 0); for the second intensity group (2- 4), they were (4.9, 2.4, .6); for the third group (5- 7), they were (9.1, 5.5, 2.1); and for the highest intensity group (8- 10), they were (13.4, 10.1, 6.5). Expressed as percentages of all surprise expressions shown, this amounts to a proportion of 83: 17: 0 in the lowest surprise group; 62: 31: 7 in the second group; 55: 33: 12 in the third, and 45: 33: 22 in the fourth. (Similar results were obtained for the subsample.) A χ^2-test of the differences in the frequency distributions of 0- 3 component displays between intensity levels (ignoring, for lack of an alternative test, the partial dependencies of the data points) was highly significant [X^2 (9,*N*= 1575)= 166.7, *P* < .001]; a parallel X^2-test including only 1- 3 component displays was marginally significant [X^2 (6, *N* = 182) = 12.3, *P* < .06]. Hence, in contrast to the earlier betweencomponents test of the intensity-coherence hypothesis, this time there was support for the hypothesis: not only was there an increase in the probability of facial expression with higher surprise intensity (cf. the results for the combined index in Table 2), but also an increase in the proportion of multiple-component responses.

DISCUSSION

The main results of the present study can be summarised as follows.

1. Replicating previous research (e.g. Lazarus et al., 1963; Ruch, 1995), within-subjects correlational analyses yielded higher coefficients and fewer negative coefficients than inter-individual analyses (for more details on this point, see Reisenzein, 1999). It may therefore be concluded that absent, low, and negative correlations between the components of the surprise syndrome obtained in previous studies (cf. the Introduction) were in part artefacts due to the use of a between-subjects design. However, other factors, such as a delayed timing of self-reports, may also have been responsible in some cases (see Schützwohl, 1998).
2. Even with an optimal data analysis design (raw data, within-subjects), the average linear correlations between the different surprise components were- with the important exception of the correlation between unexpectedness and surprise (*r* = .78)- only low to moderate, ranging from .19 (RTexpression) to .46 (surprise feeling-expression).

At this point, I need to come back to the methodological decision to exclude 13 low-expressive participants (37% of the sample) from the analyses involving the facial surprise index. As said, this decision was made to maximise the comparability of the present findings to those of Ruch (1995), although it could be independently justified by statistical considerations (lack of variance of the expression index). Nevertheless, it should be recalled that there was no evidence from the three remaining surprise components that the facially nonreactive participants were less surprised. By excluding these participants from the data analysis involving facial expression, they were implicitly treated as if they had shown the same correlations of facial to other surprise components as the average of the included subjects. In fact, however, their correlations were consistently lower (zero for those eight participants who showed no expression).

3. There was some evidence for nonlinear relations among the syndrome components, but correcting for nonlinearity did not make a great difference for the association between the surprise components.
4. Similar conclusions can be drawn with respect to measurement error in the self-report data: although the correlation between the two mental syndrome components showed a noteworthy increase (from .78 to .91), this had next to no effect on the correlations between the mental and behavioural surprise components. At least in the present study, measurement error in self-reports was apparently not an important reason for the low correlations among the mental and behavioural syndrome components.
5. Associations were also only moderate between the different components of the facial expression of surprise (eyebrow raising, eye widening, mouth opening). Most surprise faces were one-component displays; twocomponent displays were next frequent, and three-component displays were least frequent. This finding is in line with Carroll and Russell's (1997) data on the facial displays of surprise shown by movie actors (which are, however, not spontaneous), but at variance with Ekman and Friesen's (1975 p. 39) suggestion that ``the surprise brow is usually joined by wide-open eyes and dropped jaw' ' . Indeed, if one follows Ekman and Friesen, most cases coded as surprise expressions in the present study do not qualify as such, because according to these authors, when ``the surprise brow . . . appears in an otherwise neutral face, the facial expression no

longer signifies an emotion' ' (Ekman & Friesen, 1975, p. 39).
6. There was no evidence for increased coherence of the surprise syndrome with increasing surprise intensity. However, the intensity-coherence hypothesis *was* supported for the different subcomponents of facial surprise.

As noted in the Method section, the experimental design used in the present study controls for most other method factors- in particular, problems of emotion induction and measurement- that could have been responsible for the low correlations between surprise components obtained in previous studies.[11] Therefore, the conclusion seems justified that the obtained moderate to low correlations between mental and behavioural components of the surprise syndrome, as well as between the two behavioural components (RT delay and facial expression), and between the subcomponents of facial expression, are genuine. Hence, I reach more or less the same conclusion, for the emotion of surprise, as that drawn by Lazarus and his co-workers some 30 years ago for the relation between different physiological, as well as physiological and subjective measures of ``stress'': ``it must be recognized that even if assessment were completely adequate from a methodological point of view, the relationship among different measures of stress and emotion would still be modest' ' (Weinstein et al., 1968, p. 406). However, against this earlier conclusion, a skeptic could have objected that even the best-designed studies by Lazarus et al. (e.g. Lazarus et al., 1963; Mordkoff, 1964) are subject to a number of methodological problems that could have obscured stronger syndrome coherence. Most important perhaps, these studies did not focus- on neither the independent nor the dependent variable side- on a specific emotion, but were concerned with a rather broadly conceptualised state of ``psychological stress' ' or ``emotional arousal' ' . That is, the aversive films used by Lazarus et al. seem to have elicited several emotions, including anxiety, anger, and sadness (Lazarus et al., 1962). In addition, the subjective point-to-point measures used by Mordkoff (1964) asked for ratings on Wundt's (1896) dimensions of feelings (pleasantness-unpleasantness, excitement-quiescence, and tensionrelaxation) rather than for specific emotions, and each rating referred to a substantial preceding time period, namely to the prior 25 seconds of the film segment (Mordkoff, 1964). The present study avoided these problems (cf. *Method*). Nonetheless, I still found only moderate coherence between the mental (unexpectedness, surprise) and behavioural (RT, expression) syndrome components of surprise, between the behavioural components, and between the subcomponents of facial expression.

Implications for Syndrome Views of Surprise and Emotion

Assuming that the foregoing conclusion is correct, what are its implications for the syndrome view of surprise (and perhaps for syndrome theories of emotion in general)? I believe that the two major implications are the following.
1. The finding of positive (if for the most part only modest) associations among the components of the surprise syndrome speaks against those skeptical views that deny any association between emotion syndrome components (implying that ``emotion syndromes' ' are fictions). Rather, the syndrome view of emotions does contain a kernel of truth- at least for surprise.

---

[11] A reviewer suggest ed that, despite the attempts to diminish the achievement character of the task, the discovery of not having correct ly answered a question could have caused frustration or embarrassment. A recent study using the quiz paradigm in which the participants completed a mood questionnaire before and after the quiz did not support this possibil ity; in fact, mood tended to improve. However, even if embarrassment or frustration occurred, this need not have lowered the correlations between the surprise syndrome components. Support ing this suggest ion, the correlat ions between the number of quiz items wrongly answered by a participant, and the six intra-individua l correlat ions among syndrome components, were all nonsig nificant and slightly positive (.03 to .27), suggest ing that, if anything, lower performance was associated with stronger coherence among syndrome components.

2. The magnitude of the obtained associations between mental and behavioural syndrome components is much too low to support a syndrome theory of surprise that assumes a deterministic, or even just a strong probabilistic association of the syndrome components. Rather, the only tenable version of a syndrome theory of surprise *that includes behavioural components* is a weakly probabilistic one: The surprise syndrome is a constellation of mental and behavioural components that are in part only weakly associated- particularly as regards the association between mental and behavioural, and among different behavioural components- and whose association may depend on context and participants.[12]

I anticipate that these conclusions will be readily accepted by readers for the behavioural surprise component investigated in this study (action delay), but that the second conclusion will meet with more skepticism in the case of facial expression. The reason is presumably (a) that it is generally accepted, and in any case easy to make plausible, that reaction time is influenced by multiple factors, and (b) that one can concretely enumerate a number of surprise-related processes- specifically, processes concerned with the analysis and evaluation of the unexpected event- that could plausibly influence the duration of RT delay apart from degree of unexpectedness and intensity of felt surprise (cf. Meyer et al., 1997b).[13] In contrast, there is a strong tradition that links particular facial expressions rather uniquely to emotions, and it may also seem more difficult at first to think of processes that could influence the facial surprise display apart from unexpectedness and felt surprise (and display rules). But neither of these considerations means, of course, that such additional influences on the surprise display, that weaken its relation to other syndrome components, do not exist. In fact, even Darwin (1872/1965) assumed that the *immediate* mental cause of brow raising and eye widening in response to unexpected events was at least originally (i.e. in our ancestors), and perhaps is still today, neither the appraisal of unexpectedness nor the feeling of surprise, but the (surprise-caused) ``desire . . . to perceive the cause [of surprise] as quickly as possible' ' (pp. 280- 281). Alternatively, it is conceivable that the surprise display is shown only if a schema discrepancy is not resolved within a certain time, or when the rapid investigation of the unexpected event requires vigorous visual search (Blurton Jones & Konner, 1971; Reisenzein & Bördgen, 1998; Reisenzein et al., 1996). The present data do not speak directly to these theoretical possibilities. However, they suggest that investigating these and related hypotheses may be more fruitful than to continue pursuing exclusively the traditional ways of accounting for dissociations between facial displays and other syndrome components- namely, trying to explain them away by attributing them to methodological artefacts, or to attempts to suppress or mask the displays (e.g. Ekman, Friesen, & Ellsworth, 1982).

---

12 These conclusions are, of course, based on the surprise syndrome component s measured in the present study; other presumably important component s, in particular physiological ones, were not assessed. However, at least for SCR, the strength of association between this variable and other syndrome component s that could be obtained in the present paradigm can be estimated from (a) the associations between SCR, RT, and surprise ratings found in previous studies (cf. the Introduction) and (b) the relative increase in the magnitude of the correlations obtained by moving from the inter-individual design of these studies to the intraindividual design used here (cf. Reisenzein, 1999). For example, assuming an inter-individual correlation of .30 between SCR and subjective measures (Niepel, 1996), an intra-individual correlation of about .45 can be expected (Reisenzein, 1999). Thus, it seems unlikely that SCR would show stronger associat ions to other component s of the surprise syndrome than the behavioural and expressive components measured in the present study. Even if this were the case, the conclusions drawn for RT delay and facial expression would be unchanged: At least with respect to these syndrome components, only a weakly probabilistic syndrome concept is tenable.

13 However, if one accepts this reasoning for action delay, one should also be prepared to accept it for peripheral-physiological response s and for goal-directed actions, for the situation is much the same for these response systems. This constitutes additional, theoretical support for the proposed empiric al extrapolation of previous and present results to the likely strength of association of SCR with the remaining surprise components (note 12).

The method artefact explanation has already been discussed. As to the hypothesis that the surprise displays in the present study were suppressed or masked, this hypothesis faces the following problems. First, it is not selfevident to me that there exist widely shared display rules that demand to suppress or mask one's surprise display in response to unexpected quiz item solutions in a nondemanding and nonthreatening situation like that staged in the present experiment, where participants were told that the answers to some items were unknown to many people. Second, the experimental situation was minimally social in character, which should have reduced the effects of any existing display rules. Supporting this expectation, Ruch (1995) found no evidence for a suppression or masking of smiling and laughter (plus high correlations between facial expression and funniness ratings) in a similar, minimally social situation. Third, even assuming that the display rule hypothesis can account for the complete or near-complete absence of surprise displays in many participants, it seems to me that it cannot *per se* (once nonlinearity has been controlled for) explain the obtained low *correlations* between surprise displays and other syndrome components. This would require additional assumptions, such as that participants randomly controlled their surprise displays to some unexpected item solutions but not to others. Fourth, a re-examination of the tapes of low-expressive participants revealed no signs of attempts to suppress or mask an incipient facial surprise display (e.g. by pressing the lips together, by shutting the eyelids, or by deliberate frowning). Fifth, two recent studies using, respectively, the quiz paradigm (Reisenzein & Mondkowsky, in prep.) and a different surprise paradigm (Reisenzein & Bördgen, 1998, exp. 3) experimentally manipulated the level of sociality, but failed to find significant effects on facial surprise.

It might also be argued that the main reason for the low association of facial surprise to the other components of the surprise syndrome was that- notwithstanding the self-ratings of surprise reported in Table 1- the participants were only very mildly surprised even by highly unexpected quiz solutions, and that higher correlations might have been obtained had stronger surprise been induced. In a closely related vein, it could be argued that higher correlations might have been obtained if facial displays that were too weak to be visible on the videotapes had been measured via EMG (cf. Tassinary & Cacioppo, 1992). These possibilities can be definitely answered only by future research. For the time being, however, I would like to point out three problems for the ``insufficient intensity hypothesis' ' . First, as noted, this hypothesis is forced to reject, or at least to reinterpret, the self-reports of surprise. It should be noted in this context that the maximum incidence of surprise displays obtained for a particular item reported in Table 1 (35%) did not increase if only the two highest categories (9- 10) of the surprise ratings were considered (33% expression), or if only items with surprise ratings *and* RT delays from the upper 10% of the respective distributions were taken into account (30% expression). Second, although the insufficient intensity hypothesis could explain a low *incidence* of surprise displays, similar to the display rule hypothesis it cannot *per se* explain the low intra-individual *correlations* of facial expression with the remaining syndrome components. For this, one must additionally assume that syndrome coherence increases with increasing surprise intensity- a hypothesis that, as reported, was not supported by the data. Third, a study by Reisenzein and Bördgen (1998, exp. 2) in which, according to the participants' self-ratings, higher average surprise was induced by an unexpected event than by the most surprising quiz item used in the present studies, failed to obtain a higher incidence of surprise displays. On the contrary, only 9% of the participants showed at least one component of the surprise expression in this study.

A Central State Version of Syndrome Theory

Some readers may feel that the proposed, weakly probabilistic concept of ``surprise syndrome' ' amounts in effect to an abandonment of the syndrome theory of surprise *as originally conceived* (particularly if- as the weakly probabilistic view of syndromes almost inevitably requires- one also allows for context- and person-dependence of the association among the syndrome components). Whereas this conclusion is certainly true for deterministic versions of syndrome theory, it is more difficult to evaluate for its probabilistic variants, because most adherents of probabilistic syndrome theories of emotions have not stated precisely how weak an association between syndrome components they still deem acceptable. The most extreme position that one *could* take on this issue is to require no more than that, pooled across situations and individuals, the correlation between syndrome components is > 0. However, I doubt that any of the syndrome theorists quoted in the introduction would feel comfortable with such a weak version of syndrome theory (cf. Lang, 1988).

However this may be, there is a possibility of holding on to a stronger concept of ``surprise syndrome' ' even if the present findings are accepted as valid. This can be achieved by restricting the components of the surprise syndrome to a subset of the mental (or central-neurophysiological) processes associated with surprise. Although the obtained correlations between mental and behavioural syndrome components were only moderate, they are strong enough to support the hypothesis that there exist strong associations among underlying, central processes. This ``central state' ' version of syndrome theory can be obtained from the behavioural version considered so far by changing the latter theory in two respects. (1) The concept of ``response delay' ' or ``action interruption' ' is reinterpreted to refer to but a *component* of the observable delay, for example to the time interval between the registering of unexpectedness and the resumption of the interrupted processing of the parallel task; or even more narrowly, to a subphase of this interval, for example the duration of the initial, ``shock' ' phase of the surprise process (e.g. Shand, 1920). (2) Analogously, in place of observable facial expression, central state syndrome theory posits a central process that controls facial activity. This could either be the activity of a ``central motor mechanism' ' (e.g. Leventhal, 1984) or even better, a mental state that controls this mechanism (i.e. an ``expressive impulse' '). Of decisive importance in this context would be the assumption that, even in the absence of attempts to control facial expression, the posited expressive impulse need not cause facial muscle activity, but has this effect only if additional contextual or other conditions are fulfilled (e.g. Reisenzein et al. 1996). A consequence of this reinterpretation of the behavioural and expressive components of the surprise syndrome is, of course, that the surprise syndrome becomes a purely mental (or, neurophysiologically speaking, a central-physiological) constellation of processes. More precisely, the central state surprise syndrome is a pattern of covarying activity in certain central nervous system modules that can manifest itself only weakly in observable behaviour, and need not show in behaviour at all. (This theory can of course be extended to other emotion syndromes, such as fear or anger.)

Although the data reported in this article are insufficient to establish the proposed central state syndrome theory of surprise, they are consistent with this theory, and are more easily explained by it than by the behavioural version of syndrome theory. Hence, central state syndrome theory recommends itself as an ``inference to the best explanation' ' (cf. Harman, 1989). In addition, the central state version of syndrome theory is recommended by two theoretical considerations. First, it seems to me that the proposal to include observable behaviours as literal components into emotional syndromes- as opposed to treating them as mere fallible indicators of central emotion states- was originally motivated, at least in part, by a behaviouristic philosophy of psychology that is now widely regarded as obsolete. Second, the central state version of emotional syndrome theory is suggested by a psychoevolutionary perspective on emotions that focuses on the phylogenetic development of psychological mechanisms underlying behaviour, rather than on that of behaviour *per se* (Cosmides & Tooby, 1987). Although not exactly what most syndrome theorists of emotion seem to have intended, central state syndrome theory may nevertheless be welcomed by them as an honourable retreat in the face of the problems encountered by the behavioural version of syndrome theory.

Manuscript received 19 August 1998
Revised manuscript received 9 March 1999

REFERENCES

Averill, J.R. (1980). A construc tivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 1, pp. 305- 339). New York: Academic.
Barry, R.J. (1982). Novelty and significance effects in the fractionation of phasic OR measures: A synthesis with traditional OR theory. *Psychophysiology*, *19*, 28- 35.
Barry, R.J. (1990). The orienting response : Stimulus factors and response measure s. *Pavlovian Journal of Biological Science*, *25*, 93- 103.
Barry, R.J., & James, A.L. (1981). Fractionation of phasic respons es in a dishabituation paradigm. *Physiology and Behavior*, *26*, 69- 75.
Blurton Jones, N.G., & Konner, M.J. (1971). An experiment on eyebrow-raising and visual searching in children. *Journal of Child Psychology and Psychiatry*, *11*, 233- 240.
Bortz, J., Lienert, G.A., & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik* [Nonparametric methods in biostatistics]. Berlin: Springe r.
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 sympathet ic activation in emotion: A critical review, with emphasi s on modeling underlying mechanisms and individual differences. *Journal of Personali ty and Social Psychology*, *62*, 110- 128.
Charlesworth, W.R. (1964). Instigation and maintenance of curiosity behavior as a function of surprise versus novel and familiar stimuli. *Child Development*, *35*, 1169- 1186.
Carroll, J.B. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, *26*, 347- 372.
Carroll, J.M., & Russell, J.A. (1997). Facial expressions in Hollywood's portraya l of emotion. *Journal of Personali ty and Social Psychology*, *72*, 164- 176.
Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior*, *15*, 141- 154.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measu rement*, *20*, 37- 46.
Cosmide s, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. DupreÂ (Ed.), *The latest on the best. Essays on evolution and optimality* (Vol. 11, pp. 277- 306). Cambridge, MA: MIT Press.
Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago, IL: University of Chicago Press. (Original work published 1872)
Davidson, R.J. (1992). Prolegomina to the structure of emotion: Gleanings from neuropsychology. *Cognition and Emotion*, *6*, 245- 268.
Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571- 582.
Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (Vol. 19, pp. 207- 283). Lincoln, NE: University of Nebraska Press.
Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169- 200.
Ekman, P., & Friesen, W.V. (1975). *Unmasking the face*. Englewood Cliffs, NJ: Prentice-Hall.
Ekman, P., & Friesen, W.V. (1978). *Facial action coding system*. Palo Alto, CA: Consult ing Psychologists Press.
Ekman, P., Friesen, W.V., & Ellsworth, P. (1982). Does the face provide accurate information? In P. Ekman (Ed.), *Emotion in the human face* (2nd ed., pp. 56- 110). Hillsdale, NJ: Erlbaum.
Fazio, R.H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendric k & M.S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74- 97). Newbury Park, CA: Sage.
Frijda, N.H., Ortony, A., Sonnemans, J., & Clore, G.L. (1992). The complexity of intensity: Issues concerning the structure of emotion intens ity. In M.S. Clark (Ed.), *Review of Person ality and Social Psychology* (Vol. 13, pp. 60- 89). Newbury Park, CA: Sage.
Fridlund, A.J. (1994). *Human facial expre ssio n: An evolutionary view.* San Diego, CA: Academic Press.
Frois-Wittmann , J. (1930). The judgment of facial expre ssion. *Journal of Experimental Psychology*, *8*, 113- 151.
Green, D.P., Goldman, S.L., & Salovey, P. (1993). Measurement error masks bipolari ty in affect rating s. *Journal of Personali ty and Social Psychology*, *64*, 1029- 1041.
Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, *22*, 38- 77.
Harman, G. (1989). The inference to the best explanation. In R. Brody & R. Grandy (Eds.), *Readings in the philosophy of science* (pp. 323- 328). Englewood Cliffs, NJ: Prentice-Hall.
Hodgson, R., & Rachman, S. (1974). Desynchrony in measures of fear. *Behaviour Research and Therapy*, *12*, 319- 326.
Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personal ity and Social Psychology*, *69*, 280- 288.
Hiatt, S.W., Campos, J.J., & Emde, R.N. (1979). Facial patterning and infant emotional expression: Happine ss, surprise, and fear. *Child Development*, *50*, 1020- 1035.
Izard, C.E. (1977). *Human emotions*. New York: Plenum.
Karabinus, R.A. (1975). The *r*-point biserial limitation. *Education and Psychological Measurement*, *35*, 277- 282.
Landis, C. (1924). Studie s of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology*, *4*, 447- 509.
Lang, P.J. (1988). What are the data of emotion? In V. Hamilton, G.H.Bower, & N.J. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 173- 191). Dortrecht : Kluwer.
Lazarus, R.S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
Lazarus, R.S., Averill, J.R., & Opton, E.M.jr. (1970). Toward a cognitive theory of emotion.

In M.B. Arnold (Ed.), *Feelings and emotions* (pp. 207- 232). New York: Academic Press.
Lazarus, R.S., Speisman, J.C., Mordkoff, A.M., & Davison, L.A. (1962). A laboratory study of psychological stress produced by a motion picture film. *Psychological Monographs*, *76* (whole No. 553).
Lazarus, R.S., Speisman, J.C., & Mordkoff, A.M. (1963). The relationship between autonomic indicators of psychological stress: Heart rate and skin conductance. *Psychosomatic Medicine*, *25*, 19- 30.
Lazarus, R.S., Tomita, M., Opton, E.M., Jr., & Kodama, M. (1966). A cross-cul tural study of stress reaction patterns in Japan. *Journal of Personali ty and Social Psychology*, *4*, 622- 633.
Leventhal, H. (1984). A perceptual-motor theory of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 117- 182). New York: Academic Press.
McDougall, W. (1960). *An introduction to social psychology*. London: Methuen. (Original work published 1908).
McGhee, P.E. (1977). Children's humour: A review of curren t re search trends. In A.J. Chapman & H.C. Food (Eds.), *It' s a funny thing, humour* (pp. 199- 209). Oxford: Pergamon.
Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion*, *5*, 295- 311.
Meyer, W.-U., & Niepel, M. (1994). Surprise. In V.S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 4, 353- 358). Orlando, FL: Academic Press.
Meyer, W.-U., Niepel, M., & Schützwohl, A. (1994). UÈ
berraschung und Attribution [Surprise
and attribution]. In F. Förste rling & J. Stiensmeier-Pelster (Eds.), *Attributionstheorie, Grund lagen und Anwendungen* (pp. 105- 121). Göttingen: Hogrefe.
Meyer, W.-U., Schützwohl, A., & Reisenz ein, R. (1997a). *Einführung in die Emotionspsychologie: II: Evolutionspsycho logische Emotionstheorien* [Introduc tion to the psychology of emotions: *II*: Evolutionary -psychological theories]. Bern; Huber.
Meyer, W.-U., Reisenz ein, R., & Schützwohl, A. (1997b). Towa rds a proc ess analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*, 251- 274.
Mondkowsky, I. (1996). Effekte des sozialen Kontextes auf den Ausdruck von UÈ
berraschung
[Effects of social context on the express ion of surprise]. Unpubli shed diploma thesis, Universi ty of Bielefeld.
Mordkoff, A.M. (1964). The relationship between psycholo gical and physiolog ical response to stress. *Psychosomatic Medicine*, *26*, 135- 150.
Neter, J., & Wasserman, W. (1974). *Applied linear statistical models*. Homewood , IL: Irwin.
Niepel, M. (1996). *Reaktionen auf unerwartete Ereignisse* [Reactions to unexpected events]. Unpubli shed dissertation, University of Bielefeld.
Niepel, M., Rudolph, U., Schützwohl, A., & Meyer, W.-U., (1994). Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. *Cognition and Emotion*, *8*, 433- 452.
Opton, E.M., Jr., Rankin, N.O., & Lazarus, R.S. (1965). A simplified method of heart rate measurement. *Psychophysiology*, *2*, 87- 97.
Parrott, W.G., & Gleitman, H. (1989). Infants' expectations in play: The joy of peek-a-boo. *Cognition and Emotion*, *3*, 291- 311.
Plutchik, R. (1980). *Emotion. A psychoevolutionary synthesis*. New York: Harper & Row.
Rachman, S., & Hodgson , R.I. (1974). Synchrony and desynchrony in fear and avoidance. *Behaviour Research and Therapy*, *12*, 311- 318.
Reisenzein, R. (1999). *Effects of data analysis strategy on the strength of association between components of the surprise syndrome*. Manuscript, University of Bielefeld.
Reisenzein, R., & Bördgen, S. (1998). *Evidence for strong dissociation between emotion and facial displays: The case of su rp rise*. Manuscript, University of Bielefeld.
Reisenzein, R., Meyer, W.-U., & Schützwohl, A. (1996). Reactions to surprising events: A paradigm for emotion research. In N. Frijda (Ed.), *Proceedings of the 9th conference of the In ternational Society for Research on Emotions* (pp. 292- 296). Toronto: ISRE.
Reisenzein, R., & Mondkowsky, I. (in prep.). *Effects of sociality and emotion intensity on the nonverbal display of surprise.*
Roseman, I.J. (1996). Why the se appraisals? Anchoring appraisal mode ls to research on emotional behavior and related response systems. In Frijda, N.(Ed.), *Proceedings of the 9th conference of the International Society for Research on Emotions* (pp. 106- 110). Toronto: ISRE.
Rosenberg, E.L., & Ekman, P. (1994). Coherence between expressive and experiential sy stems in emotion. *Cognition and Emotion*, *88*, 201- 229.
Ruch, W. (1990). *Die Emotion Erheiterung: Ausdrucksformen und Bedingungen* (The emotion of exhilaration: Forms of expression and eliciting conditions]. Unpublis hed Habilitation thesis, Department of Psychology, Universi ty of Düsseldorf , Germany.
Ruch, W. (1995). Will the real relationship between facial expression and affective experience please stand up: The case of exhilaration. *Cognition and Emotion*, *9*, 33- 58.
Schmidt-Atzert, L. (1993). *Die Entstehung von Ge fühlen: Vom Auslöse r zur Mitteilung* [The generation of feelings: From eliciting stimulus to verbal report]. Berlin: Springe r.
Schützwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1182- 1199.
Schützwohl , A., & Reisenzein, R. (1998). Children's and adults ' reactions to a schemadiscrepant event: A developmental analysis of surprise. *International Journal for Behavioral Development*, *23*, 37- 62.
Scherer, K.R. (1984). On the nature and function of emotion: A component process approac h. In K.R. Schere r & P. Ekman (Eds.), *Approaches to emotion* (pp. 293- 317). Hillsdale, NJ: Erlbaum.
Shand, A.F. (1920). *The foundations of character*. London: Macmillan.
Siddle, D.A.T., & Jordan, J. (1993). Effects of intermodality change on electrodermal orienting and on the allocation of processing resource s. *Psychophysiology*, *30*, 429- 435.
Smith, C.A., & Scott, H.S. (1997). A componential approac h to the meaning of facial expression. In Russell, J.A., & FernaÂ ndez-Dols, J.(Eds.), *The psychology of facial expressio n*

(pp. 229- 254). Cambridge : Cambridge Universi ty Press.
Stiensmeier-Pelster, J., Martini, A., & Reisenz ein, R. (1995). The role of surprise in the attribution process. *Cognition and Emotion*, *9*, 5- 31.
Tassinary, L.G., & Cacioppo, J.T. (1992). Unobservable facial actions and emotion. *Psychological Science*, *3*, 28- 33.
Theios, J., & Amrhein, P.C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review*, *96*, 5- 24.
Tooby, J., & Cosmide s, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*, 375- 424.
Vasey, M.W., & Thayer, J.F. (1987). The continuing problem of false positives in repeated measure s ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, *24*, 479- 486.
Wallbott, H.G., & Ricci-Bitti, P. (1993). Decoders' processin g of emotional facial expression- a top-down or bottom-up mechanism? *European Journal of Social Psychology*, *23*, 427- 443.
Weinstein, J., Averill, J.R., Opton, E.M.,Jr., & Lazarus, R.S. (1968). Defens ive style and di screp ancy betwe en se lf-report and physio logical indexes of st re ss. *Psychosomatic Medicine*, *10*, 406- 413.
Wheldall, K., & Mittler, P. (1976). Eyebrow- raising, eye widening and visual search in nursery school children. *Journal of Child Psychology and Psychiatry*, *17*, 57- 62.
Wundt, W. (1896).*Grundriss der Psychologie* [Outlines of Psychology]. Leipzig: Engelmann.