Journal of Personality and Social Psychology
2006, Vol. 91, No. 2, 295–315

0022-3514/06/$12.00 DOI: 10.1037/0022-3514.91.2.295

Evidence for Strong Dissociation Between Emotion and Facial Displays: The Case of Surprise

Rainer Reisenzein
University of Greifswald

Sandra Bördgen and Thomas Holtbernd
University of Bielefeld

Denise Matz
University of Bochum

Eight experiments examined facial expressions of surprise in adults. Surprise was induced by disconfirming a previously established schema or expectancy. Self-reports and behavioral measures indicated the presence of surprise in most participants, but surprise expressions were observed only in 4%–25%, and most displays consisted of eyebrow raising only; the full, 3-component display was never seen. Experimental variations of surprise intensity, sociality, and duration/complexity of the surprising event did not change these results. Electromyographic measurement failed to detect notably more brow raisings and, in one study, revealed a decrease of frontalis muscle activity in the majority of the participants. Nonetheless, most participants believed that they had shown a strong surprise expression.

Keywords: emotion, facial expression, surprise

Since the publication of Darwin's (1872/1998) book on the expression of emotions in man and animals, the relation between emotional states and facial displays has been a controversial topic. Most empirical research focused on the inference of emotions from posed facial expressions (for reviews, see Elfenbein & Ambady, 2002; Russell, 1994). This research revealed substantial agreement among judges, both intra- and cross-culturally, that a few emotions—in particular, happiness, sadness, fear, anger, disgust, and surprise—are associated with specific facial displays (e.g., Ekman, Friesen, et al., 1987; for a recent example, see Tracy, Robins, & Lagattuta, 2005). Many believe that these findings constitute strong evidence for the existence of phylogenetically determined, discrete emotion mechanisms that comprise motor programs for emotion-specific facial displays as core components. Indeed, for some emotion theorists, the existence of a facial display that, unless inhibited or masked, is shown whenever one has the emotion comes close to a defining condition for emotions (e.g., Ekman, 1997; Izard, 1991; Leventhal, 1984; Tomkins, 1962).

However, in recent years, objections have been raised against this "emotions view of faces" (Fridlund, 1994, p. 124) or, as we call it here, the *affect program theory* (APT) of facial displays. First, it has been argued that the agreement among judges on the association between facial expressions and emotions is not as high as the proponents of APT have claimed (Russell, 1994; for a reply, see Ekman, 1999). Second, it has been recalled to the scientific public's mind that the facial judgment studies that constitute the central piece of evidence for APT are first and foremost studies of folk–psychological beliefs about the association between emotions and facial displays (Reisenzein, 2000a; Russell, Bachorowski, & Fernández-Dols, 2003). Although the intra- and intercultural consistency of these beliefs suggests that they contain a kernel of truth, they (or at least their usual interpretation) could be erroneous in theoretically important respects. A salient possibility is that these beliefs reflect not the modal association between emotions and facial displays but their association in *ideal-type* cases, in which all components of the emotion syndrome are present (Horstmann, 2002). These ideal types could, however, be rarely exemplified in everyday life. Supporting this possibility, research on the actual association between emotions and facial displays suggests that this association is not as strong as APT seems to imply. In particular, facial expressions of emotion are often absent in situations in which, at first sight at least, APT would predict them to occur (e.g., Fernández-Dols & Ruiz-Belda, 1997; Fischer, Manstead, & Zaalberg, 2003; Russell et al., 2003). In addition, if they occur, facial expressions of emotion seem to be more often partial than complete (e.g., Carroll & Russell, 1997; Reisenzein, 2000a). Finally, at least some facial displays, notably smiling, seem to be as strongly pulled forth by the presence of other people as by the emotional state of the person (e.g., Fridlund, 1991; Holodynski, 2004; Kraut & Johnston, 1979; Ruiz-Belda, Fernández-Dols, Carrera, & Barchard, 2003; for reviews, see Fischer et al., 2003; Parkinson, 2005; Wagner & Lee, 1999).

Rainer Reisenzein, Institute of Psychology, University of Greifswald, Greifswald, Germany; Sandra Bördgen and Thomas Holtbernd, Department of Psychology, University of Bielefeld, Bielefeld, Germany; Denise Matz, Department of Medical Psychology and Medical Sociology, University of Bochum, Bochum, Germany.

We thank Alfons Hamm and Almut Weike for their assistance with the EMG measurement in Experiment 7 and Gernot Horstmann, Michael Niepel, and Achim Schützwohl for their comments on earlier versions of the article.

Correspondence concerning this article should be addressed to Rainer Reisenzein, Institute of Psychology, University of Greifswald, Franz-Mehring-Straße 47, 17487, Greifswald, Germany. E-mail: rainer.reisenzein@uni-greifswald.de

However, the issues are clearly not yet decided. First, as discussed in more detail later, APT has conceptual resources that allow this theory to explain many cases of reported dissociations between emotions and facial displays (Ekman, 1997; Rosenberg & Ekman, 1994). Second, the available evidence suggests that emotional states affect facial displays at least in addition to other factors (e.g., Hess, Banse, & Kappas, 1995; Jakobs, Manstead, & Fischer, 2001; Reisenzein, 2000a; Ruch, 1995). Third, research has so far concentrated on the facial display of smiling and associated emotions such as happiness or amusement, whereas the relation of other basic emotions to facial displays has been much less studied (for research on disgust, see, e.g., Rosenberg & Ekman, 1994; for research on sadness, see, e.g., Jakobs et al., 2001; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; research on surprise is reviewed below). It is conceivable that APT holds true for some emotions but not for others. Therefore, before sweeping generalizations are made, a wider range of basic emotions and associated facial displays must be taken into view.

Motivated by these considerations, the aim of the studies reported here was to examine the affect program theory of facial displays, plus several modifications of this theory, for the case of surprise.[1] Surprise recommended itself as a suitable object of inquiry for several reasons. First, at least since Darwin, surprise has been associated with a biologically determined facial display consisting of eyebrow raise, widening of the eyes, and opening of the mouth/jaw drop; and since Darwin, surprise appears on the lists of most basic-emotion theorists, including all proponents of APT. Indeed, if one accepts the standard view of (biologically) basic emotions—organized response syndromes that are typically elicited by cognitive appraisals of situations, and that evolved as "solutions" of recurrent adaptive problems (e.g., Cosmides & Tooby, 2000; Ekman, 1999)—surprise seems to be as good an example of a basic emotion as one can find (see Meyer & Niepel, 1994). Second, the relation between emotion and facial expression has not yet been studied in depth for the case of surprise. Third, compared with other basic emotions such as anger or fear, the study of surprise offers a number of advantages (Reisenzein, 2000a). In particular, surprise can be easily induced in laboratory settings with excellent control of its onset, intensity, and the timing of its measurement. In addition, apart from subjective reports of surprise, a variety of other self-report and nonfacial behavioral indicators of surprise are available, and ethical problems associated with the induction of other emotions can be largely avoided. However, before we give an overview of our studies, we briefly review previous relevant research.

Review of Research

Studies With Children

The majority of the researchers who studied surprise expressions in children investigated infants. To induce surprise, they typically used a repetition–change paradigm (i.e., a salient stimulus change after a series of no-change trials), often in combination with a "magical" event. Hiatt, Campos, and Emde (1979) and Camras, Meng, Ujiie, et al. (2002) secretly switched a hidden toy that the infants previously had repeatedly retrieved; Parrott and Gleitman (1989) changed the identity or location of the experimenter in a peek-a-boo game after a series of no-change trials; and Scherer, Zentner, and Stern (2004) changed the experimenter's voice by means of digital filtering after a period of normal talking. Hiatt et al. (1979) also attempted to induce surprise by staging the instant disappearance of a musical Ferris wheel that the infants were watching, whereas Bennett, Bendersky, and Lewis (2002) and Reissland, Shepherd, and Cowie (2002) confronted infants with a jack-in-the-box. In older children, researchers sought to induce surprise by a "magical" change of the color or number of marbles after a series of no-change trials (Charlesworth, 1964) and by means of unexpected loud noises (a loud clock buzzer or a cycle horn went off during story time; Blurton Jones & Konner, 1971; Wheldall & Mittler, 1976).

In most studies, fewer than half of the children showed at least some evidence of a surprise display (i.e., at least one component of the expression) in response to the presumed unexpected event. The maximum frequency of a single surprise component reported in any one study was 60% (Hiatt et al., 1979; for eye widening and Ferris wheel); the maximum frequency of a two-component display (eyebrow raising plus eye widening) was 52% (Bennett et al., 2002). In addition, when a surprise display occurred, it was nearly always incomplete (i.e., it consisted of only one or two components of the traditional surprise expression). For example, only 7% complete surprise displays were observed in the study by Camras et al. (2002), and only 3.5% were observed in the study by Scherer et al. (2004). Furthermore, in two studies, surprise displays occurred with the same (low) frequency in the surprise trial as in a presumably neutral baseline period (Camras et al., 2002; Scherer et al., 2004); and in one study, they occurred as often to surprising events as to events that were not intended to elicit surprise (arm restraint and the approach of a stranger; Bennett et al., 2002). Finally, one study obtained suggestive evidence for the context dependence of the link between surprise and facial expression: Blurton Jones and Konner (1971) observed about 50% eyebrow raisings in response to a clock buzzer if the clock was hidden behind an object, but they observed only approximately 5% when it was in full view.

Although these results provide some support for the assumption of APT that surprise is associated with a characteristic facial display, they also suggest that this association is far from perfect. However, as mentioned earlier, APT has conceptual resources that allow this theory to explain many cases of observed emotion–face dissociations. Exploiting these resources to the maximum, proponents of APT can argue with some plausibility that the reviewed findings are inconclusive. First, at least in some studies, the surprising events may have induced other strong emotions or facial

[1] Following Darwin (1872/1998), we conceptualize surprise as a mental state or process, and speak of the surprise display as being caused by, and in this sense as *expressing*, the mental state of surprise. However, we leave it open whether the proximate mental cause of the surprise display is the feeling of surprise, the appraisal of unexpectedness that we take to be its cognitive antecedent (cf. Meyer et al., 1997), or some other surprise-related mental process or combination of processes. Surprise must be distinguished from the startle reaction elicited by sudden intense stimuli, in particular loud noises (Ekman, Friesen, & Simons, 1985; Koch, 1999). Whereas surprise is caused by the appraisal of unexpectedness, startle is a reflexlike reaction to intense sensory input.

reactions that interfered with the surprise display. Conversely, nominally unsurprising situations in which surprise displays were seen may in fact (also) have elicited surprise (e.g., Bennett et al., 2002). Second, in all studies, the facially unresponsive children may simply not have been surprised by the eliciting events, either because they failed to attend to them or because they did not appraise them as unexpected. This possibility is difficult to exclude because most studies did not include an independent indicator of surprise, and those who did found evidence for surprise only in a subsample (e.g., Camras et al., 2002; Hiatt et al., 1979). Third, again in all studies, the experimenter, parents, or peers were present as onlookers; therefore, at least in the studies with older children, surprise expressions may have been suppressed or masked in accordance with display rules (e.g., Charlesworth, 1964). Fourth, at least in infants, the evolutionary module responsible for the surprise display may not yet be fully developed (cf. Bennett et al., 2002). When taken together, these factors may well explain the observed dissociations between surprise and its facial display in the studies with children.

Studies With Adults

Studies of emotional expressions in adults allow researchers to avoid several of the problems inherent to studies with infants and small children. Nevertheless, only four studies appear to have examined adults' facial reactions to potentially surprising events, and two of these (Ekman, 1972; Ekman, Friesen, & Ancoli, 1980) did not have surprise as their main focus.

The first study was conducted by Landis (1924) as part of his pioneering if controversial research on facial expressions of emotion. Similar to the children studies reviewed earlier, only a minority of Landis's participants showed unambiguous components of the traditional surprise display in potentially surprising situations. For example, about 30% showed eyebrow raising and about 20% eye widening when a firecracker was dropped behind their chair. However, the conclusions that can be drawn from Landis's research are limited, as his study is open to most of the objections mentioned in the discussion of the children studies and suffers from additional problems, such as an inadequate measurement of facial reactions (see Ekman, Friesen, & Ellsworth, 1982; Woodworth & Schlosberg, 1954).

Ekman (1972) coded the spontaneous facial expressions of 50 participants who, while alone, watched a stressful film and a neutral film. One hundred twenty-six pure (i.e., unblended) surprise expressions were coded during the stress film and 28 during the neutral film. However, neither the number of surprising incidents shown in the films nor the participants' subjective reactions to these incidents were measured; therefore, the strength of association between surprise and facial expression is difficult to evaluate.[2] Ekman, Friesen, and Ancoli (1980) also investigated spontaneous facial expressions during a stressful film. Although retrospective ratings suggested that some surprise was experienced during the film, facial expressions of surprise were infrequent (Ekman et al., 1980, p. 1131; the exact data were not reported). However, the film used in this study seems to have primarily elicited feelings and facial reactions of disgust that may have interfered with the expression of surprise. Therefore, this study, too, does not permit to draw firm conclusions about the relation between surprise and facial expression.

In contrast to these earlier investigations, Reisenzein (2000a) focused explicitly on surprise, which in this case was induced by confronting participants with unexpected solutions to selected items in a computerized quiz. Although subjective ratings and behavioral measures (reaction time [RT] delay on a parallel task) attested to the effectiveness of the surprise induction, maximally 34% of the participants showed a facial surprise display (at least one component) to any one item. Also, the observed surprise expressions were mostly one- or two-component displays. Hence, only moderate coherence between surprise and its facial expression was again found even when many potential problems of previous studies were controlled (for details, see Reisenzein, 2000a). Two possible remaining objections against this study, however, are that the intensity of surprise elicited by the unexpected quiz solutions was frequently too low to result in a facial display and that many participants inhibited their surprise expressions as the experimenter remained in the room during the quiz.

Taken together, the available studies on surprise expressions in children and adults suggest three main conclusions. First, at least partial surprise displays in response to the theoretically predicted elicitors (unexpected events) have been observed in controlled laboratory situations. However, so far it has not been demonstrated that the traditional expression of surprise is shown in such situations by even the majority of surprised people. Second, most, if not all, observed cases of dissociation between surprise and facial expression can be attributed by proponents of APT to methodological problems (e.g., failure to induce surprise) or substantive factors (e.g., control of expression). Third, with the exception of the studies by Blurton Jones and Konner (1971) and Reisenzein (2000a), no attempt has so far been made to empirically examine the viability of these or other possible explanations for the observed dissociations between surprise and facial expression.

Aims and Overview of the Present Studies

The experimental induction and the measurement of surprise used in the present studies was based on a cognitive–psychoevolutionary model of surprise (e.g., Meyer, Reisenzein, & Schützwohl, 1997; Reisenzein, 2000b). The core of this model, which is depicted in Figure 1, concerns the mental processes elicited by (ultimately) surprising events. According to the model, these processes begin with (a) the appraisal of a cognized event as schema-discrepant or unexpected. Disconfirmation of an explicit or an implicit expectancy has been posited by practically all surprise theorists as the primary, if not the only, cognitive elicitor of surprise and is so regarded in common sense also (see Reisenzein, 2000b; Ruffman & Keenan, 1996). The appraisal of an event as unexpected then causes (b) the occurrence of a surprise feeling and, simultaneously, the interruption of ongoing information processing and the reallocation of processing resources to the unexpected event. The function of interruption and resource reallocation is to prepare the individual for (c) the subsequent analysis and evaluation of the unexpected event plus—if the results of this analysis

[2] For example, if the participants were three times surprised during the stressful film and showed a surprise display each time, the relative frequency of surprise expressions would be a high 84%. However, if they were five times surprised and showed one surprise display to a subjectively unsurprising event, the relative frequency would be a low 30%.

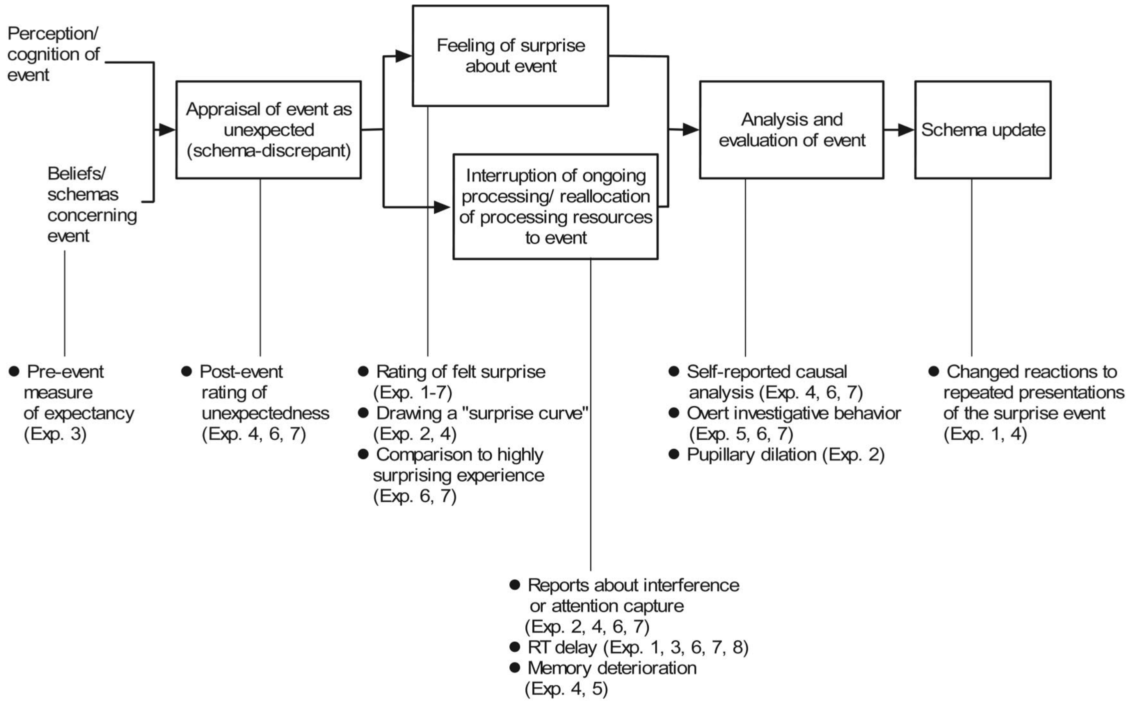

Figure 1. Model of surprise-related mental processes and their indicators in Experiments 1–8. RT = reaction time.

indicate so—the updating of the relevant schemas. The first two steps in this series of mental 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 and update the person's schemas or belief system in the face of unexpectedness (belief disconfirmation, schema discrepancy). Although this model of surprise shares an evolutionary focus with APT, it does not contain specific assumptions about the facial expression of surprise except for the assumption that this expression is caused by one or several of the processes posited in the model, possibly in conjunction with other mental events.

In accord with the described model, surprise was induced in the present studies by first establishing and then disconfirming a schema or set of beliefs concerning the experimental events. Several variants of this paradigm were used (e.g., changing the appearance of task-irrelevant stimuli in a choice reaction task after a prolonged series of no-change trials; violating a rule that a color sequence followed after participants had detected this rule). To check the effectiveness of the surprise induction, we sought to verify the occurrence of the surprise-related mental processes postulated in the model. These processes were inferred, in different studies, as shown in Figure 1.

Experiment 1: Taking a First Look

The experimental paradigm used in Experiment 1 was based on the work of Meyer, Niepel, Rudolph, and Schützwohl (1991). Participants worked on a choice reaction task. After a series of uniform trials that served to establish a schema for the experimental events, there was an unannounced change of appearance of the distractor stimuli intended to induce surprise. The effectiveness of the surprise induction was checked through self-reports of surprise and RT delay at the choice reaction task. The design of the experiment was a 5 (extremity of the unexpected stimulus change) × 2 (first vs. second presentation of the stimulus change) factorial with repeated measures on the second factor. Both manipulations were intended to influence the intensity of surprise.

Method

Participants. Participants were 60 students (40 female, 20 male) at the University of Bielefeld whose mean age was 25.5 years. Thirty-two were introductory psychology students who participated as part of their study requirements. The rest were students of various other disciplines; some were volunteers, and some were paid (about $3).[3] Participants were randomly assigned to the five stimulus change conditions.

Procedure. The experiment was described to the participants as a visual reaction task. Stimuli were presented on a personal computer with a color monitor and a mouse. Each trial began with the presentation of a fixation point at the center of the participant's monitor, inside a 14 cm × 8 cm frame. The screen area outside the central frame (the background) was filled with a number pattern. After 1400 ms, the fixation cross was

[3] Care was taken to ensure that no participant took part in more than one study: (a) We sampled students of different academic disciplines, different year groups, and in one case a different university (Experiment 7); (b) at the end of each study, we asked the participants whether they had ever participated in a similar study; and (c) we checked the videos for double appearances. To avoid advance suspicions that the studies dealt with surprise, they were announced under varying, inconspicuous titles (e.g., as studies on "visual perception" or on "ability to concentrate").

replaced by two words, one above the other; 500 ms later, a small dot appeared for the duration of 100 ms either above the upper word or below the lower word. The participants' task was to react as quickly as possible to the position of the dot by pressing the right or left mouse button. The button press caused the words to disappear, thereby ending the trial. Trials were separated by 1000 ms. RT was recorded by the computer.

In Trial 38, the first surprise trial, a salient change in the appearance of the display occurred. Depending on experimental condition, this change consisted of (a) the display of one of the two words in reverse video (RV) mode, that is, as white letters against a black background; (b) a slight change of the background pattern; (c) a pronounced change of the background (broad black and white stripes); or (d) the combination of the RV display of one word with either the slight or the strong background change. Previous research using similar repetition–change paradigms found that such stimulus changes cause moderate feelings of surprise, as well as an RT delay, in most participants (e.g., Meyer et al., 1991; Schützwohl, 1998). We expected, furthermore, that at least the extremes of the stimulus changes would differ in surprisingness (cf. Schützwohl, 1998).

Immediately after Trial 38, the participants were interviewed about the nature of the stimulus changes, after which they were asked "Did the stimulus changes occurring in the last trial surprise you? If yes, how strongly?" Answers were given on a rating scale ranging from 0 (*not at all surprised*) to 10 (*as surprised as one can be*). The experiment then continued for nine more standard trials that were followed, in Trial 48, by a repetition of the surprise event.

Once the trials had begun, the experimenter sat at a table located behind the participant and oriented 90° away, where he or she busied himself or herself with other tasks. Thus, the situation was minimally social in character ("mere presence" paradigm; Guerin, 1986). At the end of the experiment, the participants were informed about the purpose of the study and the fact that they had been videotaped, and they were asked for permission to analyze the videos. This debriefing protocol was also followed in all subsequent studies.

The participants' expressive reactions were unobtrusively videotaped from the adjoining room (see Reisenzein, 2000a, for details). The camera took a frontal picture of the participant's head and shoulders slightly from above. From the original video recordings, we constructed a master tape showing the two surprise trials and two randomly selected baseline trials. Each film clip began with the onset of the fixation cross and ended with the offset of the distractor words following the participant's reaction. The clips were coded by two independent observers who were unaware of the identity of the trials (baseline or surprise) and of the aims of the experiment. Each clip was shown twice before it was coded.

A present/absent coding scheme with eight categories was used. Four categories referred to facial or vocal surprise displays. A full facial display of surprise was defined to consist of three components (see Darwin, 1872/1998; Ekman, Friesen, & Hager, 2002): raising of the eyebrows (action units [AUs] AU1–AU2 of Ekman et al.'s [2002] *Facial Action Coding System*), widening of the eyes (accomplished by raising of the upper eyelid; AU5), and jaw drop/opening of the mouth (AUs 26–27). Each facial expression component was coded as present if it occurred within the coding time window (on average 2.7 s in the first critical trial), regardless of how intense it was or how long it lasted (e.g., even the tiniest and briefest upward movement of an eyebrow was coded as "eyebrow raising present"). A fourth coding category captured surprise vocalizations such as "oh" or "wow." Mouth opening was coded in conjunction with a surprise vocalization only if it preceded the vocalization and, thus, was apparently not simply the by-product of the vocalization. The remaining three coding categories comprised smiling (AU12) and laughter, as well as various nonverbal and verbal responses reflecting cognitive reactions to the surprising event: nodding, verbal acknowledgment ("aha"), and affirmative vocalizations (e.g., "see!"). These additional categories were suggested by Reisenzein's (2000a) study.

We trained the coders using written descriptions of the coding categories, pictures of prototypical surprise expressions, and videotapes of comparatively expressive participants from the previous study by Reisenzein (2000a).

Results

To reduce the effect of outliers, we fixed RTs >2000 ms at this value (Fazio, 1990; Ratcliff, 1993; see also Meyer et al., 1991). RTs from false responses (12% in Trial 38 and 10% in Trial 48) were retained in the analyses. To obtain a baseline RT for each participant, we averaged the RTs of the nine trials immediately preceding Trial 38 and the nine trials between Trials 38 and 48.

Surprise feelings and RTs. In Trial 38, all but one participant reported at least minimal surprise (rating >0) about the stimulus change ($M = 5.4$, $SD = 2.3$). Also, all but one of the participants showed an RT increase from his or her individual baseline; for 68%, it exceeded 2.5 standard deviations. Mean RT increase from baseline was statistically significant both overall ($M = 435$ ms, $SD = 434$), $t(59) = 7.75$, $p < .001$; and for each of the five experimental groups considered separately, with ts$(11) > 3.05$, ps at least $< .02$.

In Trial 48, when the stimulus changes occurred for the second time, 53 of the 60 participants still felt minimally surprised (rating > 0) although, as predicted, less so than in Trial 38 ($M = 3.4$, $SD = 2.5$), $t(59) = 5.2$, $p < .001$. RTs in Trial 48 were still significantly above baseline overall ($M = 376$ ms, $SD = 469$), $t(59) = 6.20$, $p < .001$, and within each experimental group (p at least $< .05$, one-tailed).[4] In four of the five groups (all but the "weak background change only" group), the RT increase in Trial 48 was less pronounced than in Trial 38, as predicted. When we excluded the exceptional group, the RT reduction from Trial 38 to Trial 48 became significant ($M = -188$ ms), $t(47) = 2.9$, $p < .01$.[5]

We had also predicted differences in surprise intensity between the five experimental conditions (stimulus changes), at least between the extreme groups. In Trial 38, the obtained pattern of means conformed to predictions: Surprise ratings were lowest for the "weak background change only" condition ($M = 4.5$), and highest for the two background change-plus-RV word display conditions (both 5.9), with the other two conditions lying in between (5.3, 5.4). A comparison of the extreme groups confirmed the a priori hypothesis, $t(34) = 1.8$, $p < .05$ (one-tailed). The RT delays (increases from baseline) showed a parallel pattern, and again the comparison of the extreme groups was significant, $t(34) = 2.1$, $p < .05$ (one-tailed). In Trial 48, the differences between experimental groups in self-reports of surprise and (if the

[4] Guided by the principle that statistical tests should fit research hypotheses as closely as possible, we report test results as significant if they meet the 5% criterion for a two-tailed test or at least for a one-tailed test in the case of a priori, directed hypotheses (see Furr & Rosenthal, 2003). One-tailed tests are marked as such.

[5] The "weak background change only" group showed an unexpected RT increase compared with Trial 38, $t(11) = -2.9$, $p < .05$. Additional analyses suggested that most participants in this group could not identify the exact nature of the background change when it first occurred and therefore delayed their button press (which made the stimuli disappear) in Trial 48 to take a closer look.

"weak background change only" group is excluded) in RT delay were no longer significant.

Facial displays. Because of equipment breakdown, the video recordings of three participants were missing. Of the remaining 57 participants, 3 (5.3%) showed one component of the facial surprise display (two brow raises, one eye widening) in the first surprise trial. No surprise expressions were coded in the second critical trial, and one (brow raising) was coded in the baseline trials. Smiling/laughter occurred once in the first and three times in the second surprise trial. Agreement between the two coders was perfect for the surprise displays; there was one disagreement for smiling. Satisfactory reliability of the coding system was also documented in Reisenzein's (2000a) study, where a higher number of facial expressions were available for reliability estimation. To make sure that the coding interval in the critical trials had not been too short, the original tape recordings were re-examined. Again the same results were obtained.

Discussion

The central finding of Experiment 1 is the extremely low incidence of facial displays of surprise, coupled with evidence for the presence of surprise from self-ratings and RT delay. Judged by the latter measures, most participants were surprised about the unexpected stimulus changes in both critical trials. By contrast, only 5% of the participants showed evidence of a surprise expression in Trial 38, and all of them showed only a single component of the expression. In addition, the manipulation of surprise intensity (achieved by varying the extremity of the stimulus changes and by repeating the stimulus changes) affected self-reports of surprise and RT delay in the predicted manner, but it did not affect facial expression.

Experiment 2: Taking a Closer Look

Experiment 2 was part of a series of studies conducted to test the hypothesis that the intensity of felt surprise is influenced by the degree of mental interference caused by an unexpected event (see Reisenzein, 2000b). This experiment provided a convenient opportunity to verify the findings of Experiment 1 using (a) a somewhat different surprise paradigm that included an alternative manipulation of the intensity of experienced surprise; (b) several additional measures of the subjective experience of surprising events; and (c) a physiological indicator of surprise, pupillary dilation. Pupillary dilation is commonly listed as a component of the physiological orienting reaction that is held to be elicited by novel, unexpected, and significant events (e.g., Rohrbaugh, 1984), and it has been empirically demonstrated to occur in response to unexpected stimulus changes (e.g., Maher & Furedy, 1979). Because pupillary dilation is also a sensitive indicator of processing load or mental effort (Beatty & Lucero-Wagoner, 2000), we hypothesized that its occurrence in response to unexpected events reflects (directly) the exploration and analysis of these events (Meyer et al., 1997; see also Experiment 5). Klix, van der Meer, and Preuß (1984) even proposed that the dilation of the pupil in response to mental load may reflect an evolutionary adaptation to surprising events that occurred in the twilight, where pupillary dilation can presumably improve vision. Pupil size was determined in Experiment 2 from video recordings of one eye. The magnified recording of the eye region needed for this purpose simultaneously allowed (d) an excellent observation of even slight movements of the eyebrow and eyelid. In this way, Experiment 2 provided for an additional methodological control of the results of Experiment 1.

Method

Participants. The final sample consisted of 25 students (13 female, 12 male) from the same pool as in Experiment 1, who were between 20 and 28 years old. Four additional participants were excluded from the data analyses, three because of missing data (e.g., too-dark eyes), and one because she did not comply with the instructions. The participants were randomly assigned to one of two experimental conditions (high vs. low task difficulty).

Procedure. The experiment was described as a study on pupil size changes during text comprehension. The participant was asked to put his or her head on a chin rest in front of a computer monitor and to read a text (the beginning of a novel) presented in three-word chunks inside a small window. There were two experimental conditions differing in task difficulty. The task difficulty manipulation was intended as an alternative method to influence the intensity of felt surprise, as we previously found that an unexpected event is experienced as more surprising if it interferes more strongly with ongoing activities (see Reisenzein, 2000b). In the difficult-task condition, the text was presented quickly (each three-word chunk for 300 ms, followed by a 50-ms pause), whereas in the easy task condition, it was presented slowly (each chunk for 900 ms, followed by a 300-ms pause). The potentially surprising event occurred immediately after the last word of the text. It consisted of a computer-generated random tone sequence of 2.5-s duration and the simultaneous appearance of a meaningless sequence of ASCII signs within the text window, after which the program abruptly terminated. Thus, the overall appearance of the stimulus changes was rather like that of a computer breakdown. A few seconds later, the participants were presented with a set of index cards, each of which showed a question and a corresponding rating scale. The questions asked, among other things, for felt surprise and for the degree of interference and distraction caused by the surprising event, and they were answered on scales ranging from 0 (*not at all*) to 100 (*extremely*). In addition, as an alternative measurement of experienced surprise, we asked the participants to graph the time course of their surprise feelings on a Time (from the onset of the surprise event until 4.7 s later in 100 ms) $\times$ Intensity (0 [not at all surprised] to 100 [extremely surprised]) grid sheet (cf. Sonnemans & Frijda, 1994). These "surprise curves" were later digitized with the help of a graphic tablet.

During the experiment, the right eye of the participant including the eyebrow region was filmed with a video camera fitted with a strong zoom lens. This allowed us to record pupil size changes as well as movements of the brow and eyelid. Other facial movements were not considered in this study. The video recordings were later manually scored for pupil size changes in 400-ms intervals at several points before and until 2.4 s after the onset of the surprise event. To this end, we replayed the video on a large TV monitor and froze the picture at each measurement point. We then fitted transparencies with different-sized measurement rings (1 mm apart in diameter) over the enlarged pupil until a best fit was achieved, and we recorded the corresponding diameter. If a blink occurred at the selected measurement point, we scored the nearest previous or following frame showing a clear image of the pupil. To estimate the reliability of the measurements, a second coder repeated the measurements for three participants. The agreement of the measurements was nearly perfect, interrater $r = .96$.

Eye widening and brow raising were coded by a different observer in the interval beginning with the appearance of the tone and ending with the beginning of the ratings. Slow-motion display and repeated viewing were used. In view of the finding in Experiment 1 that only one instance of brow raising was observed during the two coded baseline trials, only the facial reactions to the surprise event were coded.

Results

Feelings of surprise and interference. According to the direct surprise ratings, 24 of the 25 participants were at least minimally surprised (rating >0; $M = 38$, $SD = 22$, $Mdn = 30$). Similarly, the average of the individual maxima of the Time × Intensity "surprise curves" drawn by the participants was 47 on the 100-point scale ($SD = 23$, $Mdn = 49$). Substantial correlations of the direct surprise ratings with key parameters of the graphs (up to $r = .88$, obtained for the individual maxima) suggest that the measurement of surprise was reliable. Further confirming the effectiveness of the surprise induction on the experiential level, 19 (76%) of the 25 participants said that the surprise event distracted them at least minimally (rating >0), and 16 (64%) said the event interfered at least minimally with the processing of the text.

Pupillary dilation. The mean of the measurements during the text reading and immediately before the surprise event served as a baseline for each participant. The pupil began to dilate after the surprise event, reaching its maximum 400–800 ms after the beginning of the tone sequence, after which it declined again and reached near-baseline at the last measurement point (2.4 s later). Paired t test ($df = 24$) comparisons with the baseline showed that the increase in pupil size was significant (ts at least >1.9; ps at least $<.05$, one-tailed), with the exception of the last two measurement points. All participants showed some degree of pupil dilation in response to the surprise event, and 11 (44%) showed an increase >2.5 standard deviations from their individual baseline.

Facial displays. No eyebrow raises were observed in response to the surprise event, but 1 of the 25 participants showed eye widening.

Effects of the experimental manipulation. In accord with the hypothesis that the intensity of felt surprise is influenced by the degree of interference caused by an unexpected event, participants in the difficult-task condition felt more surprised than those in the easy-task condition. We obtained this effect for both the direct surprise ratings, $t(23) = 1.9$, $p < .05$ (one-tailed), and the surprise graphs, which were significantly higher in the difficult-task group from the second to the seventh measurement interval, $t\text{s}(23) > 2.6$, $p\text{s} < .05$. The participants also felt more strongly interrupted by the surprise event in the difficult task condition, $t(23) = 1.8$, $p < .05$ (one-tailed). In contrast, the single case of eye widening was observed in the easy-task condition. The experimental manipulation also had no significant effects on pupil dilation in response to the unexpected event. This result is in line with the suggestion, made in the introduction to this study, that pupillary dilation mainly reflects processes concerned with the analysis of the unexpected event; for presumably these processes were similarly demanding in both experimental conditions.

Discussion

Experiment 2 replicated the findings of Experiment 1 and added to them in three ways. First, the results of Experiment 1 were confirmed for a somewhat different surprise paradigm that included a different manipulation of surprise intensity. Second, we confirmed the effectiveness of the surprise induction using a different method of measuring felt surprise (the drawing of a surprise curve), ratings of experienced interference and distraction, and pupillary dilation. Third, because of the magnified recording of the eye region, even very slight facial movements could be detected, ruling out another possible source of bias in Experiment 1.

In contrast to some of the earlier studies (cf. the introduction), the strong dissociation between surprise and facial expression observed in Experiments 1 and 2 cannot be plausibly attributed to interfering muscular movements (practically none occurred), suboptimal recording or coding of (visible) facial expressions (see Experiment 2 in particular), or obvious measurement problems concerning the indicators of surprise (e.g., failure to relate the self-report questions to a clearly specified event, or substantially delayed self-reports). Nor do the findings seem to be attributable to a failure to induce surprise in many participants: Clear instances of the theoretically posited elicitors of surprise (unexpected events) were focally presented, and both self-reports and behavioral measures indicated that the surprise induction was successful for most participants. Hence, purely *methodological* explanations of the failure to observe a (visible) surprise display in most participants of Experiments 1 and 2 seemed implausible.[6] In Experiment 3, we therefore examined in more detail the two *substantive* explanations of the findings of Experiments 1 and 2 available to APT: inhibition of facial displays and insufficient surprise intensity.

Experiment 3: Reducing Sociality, Increasing Intensity

The first substantive explanation that can be offered by APT for the findings of Experiments 1 and 2 is that motor-expressive tendencies (signals to the face; Ekman, 1997) were elicited but were suppressed in an attempt to conform to internalized social or personal norms concerning appropriate expressive behavior (i.e., display rules). We tested this hypothesis in Study 3 by leaving half of the participants alone during the experiment. Although this manipulation of sociality still fails to rule out the presence of an "implicit audience" (e.g., Chovil, 1991; Fridlund, 1994), this is not required for a test of the display-rules hypothesis. That is, to test this hypothesis, one need not assume that display rules have no effect in solitary situations; only that their effect is reduced. APT predicts this to occur because the stimuli that presumably activate display rules (other people) are not present in solitary situations or are present only in symbolic form (see also Chovil, 1991).

The second substantive explanation of the low incidence of surprise displays observed in Experiments 1–2 available to APT is this: A surprise display occurs only if surprise exceeds a certain threshold of intensity, and the facially nonreactive participants of Experiments 1 and 2 were not surprised enough (cf. Ekman, 1997; Tassinary & Cacioppo, 1992). This *insufficient-intensity hypothesis* covers two possibilities (see also Larsen, Norris, & Cacioppo, 2003): (a) The intensity of surprise was too low to cause a *visible* display, although invisible muscle changes did occur; (b) surprise

[6] To examine the possibility that the self-reports of surprise were influenced by experimental demand characteristics (Orne, 1962), we conducted two additional studies (details are available from Rainer Reisenzein). No evidence for demand effects was obtained.

was too weak to even elicit a motor signal to the face. To test the insufficient-intensity hypothesis, one can increase either the intensity of surprise or the sensitivity of facial measurement (by using electromyographic [EMG] recordings; cf. Tassinary & Cacioppo, 1992). In Experiment 3, we used the first method (increasing the intensity of surprise). This allowed us to test both versions of the insufficient-intensity hypothesis simultaneously: If surprise is strong enough, both the threshold of elicitation of the motor signal and the threshold of visibility of the resulting muscle movements should be exceeded. The method of surprise induction used in Experiment 3 was inspired by a study by Horstmann and Schützwohl (1998), who found that strong surprise can be induced by first making the participants believe that the stimulus events follow an invariable rule and then disconfirming this rule. This experimental paradigm also provided for an unobtrusive behavioral index of the participants' expectations of the events occurring in the upcoming trials.

Method

Participants. The final sample consisted of 22 students (12 female, 10 male) from the same pool as in Experiments 1 and 2, whose mean age was 23.2 years. The participants were randomly assigned to one of two experimental conditions (social vs. alone).

Procedure. Similar to Experiment 1, the participants worked on a choice RT task in which they had to react as quickly as possible to the position of a dot. However, in Experiment 3, the choice RT task was embedded into a second, rule-detection task: The dot appeared on the screen above or below a bar that was either red, green, blue, or yellow. The participants were told that the sequence of colors across trials followed an experimenter-specified rule and that their second task was to detect this rule. Beginning with Trial 9, they had to predict the color of the bar in the following trial. Although the rule underlying the color sequence was simple (red-green-blue-yellow), its detection was not trivial, because the colors had to be memorized and because the choice reaction task imposed an additional mental load. Pretesting suggested that 25–30 trials were needed to detect the rule. On this basis, 40 trials were presented. The first 39 trials consisted of nine repetitions of the four-color sequence plus its 10th repetition up to the next-to-last color. In Trial 40, the surprise trial, a black bar was presented instead of the yellow bar predicted by the rule. Immediately after Trial 40, the participants were asked to rate how surprised they felt about the occurrence of the black bar on the scale already used in Experiment 1.

Two experimental conditions were compared. In the (minimally) social condition ($n = 10$), the experimenter remained in the room during the experiment. In the alone condition ($n = 12$), the experimenter left the room after the first two or three trials and returned only after the surprise trial.

Results and Discussion

Rule detection (pre-event expectancies). The correct prediction of the entire color sequence immediately before the critical Trial 40 was used as the criterion for rule detection. According to this criterion, all but two 2 participants had detected the rule by Trial 39. On this basis, it can be assumed that the subsequent appearance of the black bar was rule-discrepant for nearly everybody. In addition, the black bar was also visually discrepant to the preceding colors (cf. Experiments 1–2).

Surprise feelings and RT delay. Replicating Horstmann and Schützwohl (1998), most participants rated themselves as strongly surprised by the appearance of the black bar: On the scale ranging from *not at all surprised* (0) to *as surprised as one can be* (10), all but three participants (86%) gave ratings ≥ 6, and 50% gave ratings ≥ 8 ($M = 7.0$, $SD = 2.2$). t-test comparisons showed that this mean was significantly higher than in Experiments 1 and 2 ($ps < .01$). We obtained no significant difference in felt surprise between the social group and the alone group, $t(20) < 1$. The RT data were analyzed analogously to Experiment 1. We found a significant RT increase from baseline in Trial 40 in both the social group, $t(9) = 3.7, p < .01$; and in the alone group, $t(11) = 5.2, p < .001$. All but two participants showed an RT increase from their individual baseline, and 13 (59%) showed an increase >2.5 SD ($M = 435$ ms; $SD = 434$). We found no significant difference between the experimental groups in RT increase, $t(20) = 1.1$, $p = .28$.

Facial displays. Brow raising occurred in 2 (9%) of the 22 participants (11% of those with surprise ratings $\geq$6). Both belonged to the social condition, but this difference is not significant (Fisher exact probability test). Other components of the surprise display were not observed. The obtained frequency of surprise displays is not significantly different from that in Experiment 1 (Fisher exact probability test, $p = .43$) or from that in Experiments 1 and 2 combined ($p = .39$). Two additional participants in the social condition reacted with a frown in response to the unexpected color change, possibly reflecting puzzlement (Darwin, 1872/1998).

Conclusions

Experiment 3 revealed that the incidence of surprise displays in an alone condition did not differ significantly from that in a (minimally) social condition. This finding speaks against the hypothesis that the low frequency of surprise expressions in this study, as well as in Experiments 1 and 2, was due to inhibition or masking. Experiment 3 also revealed that even an event rated as highly surprising by most participants did not result in significantly greater expressivity. This speaks against the insufficient-intensity hypothesis.

Experiment 4: Increasing the Duration and Complexity of the Surprise Event and Examining Beliefs About Expression

If the conclusions drawn from Experiment 3 are correct, then the APT of facial expressions of surprise in its original form is untenable. However, this does not mean that the more general assumption of this theory—that there is an evolutionary link between surprise and facial expression—has to be discarded as well. It is possible to expand or modify APT in ways that preserve this central idea but provide for new conceptual resources to explain the dissociation findings (whether researchers should still call these modifications "variants of APT" is another question). The general principle underlying these modifications is to make emotional displays (here the surprise display) depend on factors in addition to the presence of the emotion and the absence of deliberate control (Reisenzein, Meyer, & Schützwohl, 1996; see also Ekman, 1993, for a number of suggestions in this direction). This move necessarily weakens the link between emotion and expression, but in contrast to more radical alternatives (e.g., Fridlund, 1994), it does not completely sever this link. In Experiments 4–6, we tested several possible modifications of the APT of surprise.

In Experiment 4, we tested what is perhaps the most conservative modification of APT. Precisely speaking, this modification covers a whole set of hypotheses which have in common the assumption that they view *duration* as a critical variable for the elicitation of the surprise display: It is assumed that surprise manifests itself in the face only given some minimal duration of (a) the event that causes a schema discrepancy, (b) the schema discrepancy itself (i.e., the time needed for resolving the discrepancy and for schema update), or (c) the feeling of surprise caused by the schema discrepancy. Assuming that any one of these time spans is critical for a surprise display to occur, it may have been too brief in the preceding studies. To test this hypothesis, we simultaneously varied the duration and complexity of the surprising event.

In addition to this main goal, we had two other aims in Experiment 4. First, we wanted to provide further evidence for the presence of surprise in our participants. For this purpose, in Experiment 4 we included a fairly comprehensive set of self-report measures tapping different aspects of the subjective experience of surprising events that are predicted to occur by the model described in the introduction. Second, we queried what the participants themselves thought about their facial displays.

Method

Participants. The participants were 22 students (14 female, 8 male) from the same pool as in the previous experiments, with a mean age of 23.9 years. They were randomly assigned to one of two experimental conditions (short/simple vs. extended/complex).

Procedure. A short-term memory paradigm served as the parallel task during which participants were distracted by a surprising event. The memory task comprised 54 trials. Each trial began with the presentation of a fixation cross at the center of the computer monitor. Five hundred milliseconds later, seven different, randomly selected consonants were simultaneously shown for 4 s. This was followed by a rehearsal period of 4 s, symbolized on the monitor by a 20-step countdown that began with the number 20 and ended with a question mark in place of the zero. The numbers were presented successively at the center of the screen for 200 ms, and each number was accompanied by a brief but fairly loud tone. The participants were asked to memorize the letters and to report back as many as they could when the question mark appeared. The recalled letters were noted down by the experimenter.

There were five critical trials: Trials 20, 28, 34, 50, and 54. The potentially surprising event consisted of several of the countdown numbers being shown in reverse video mode (white against black), a one-octave pitch change of the accompanying tones, or both stimulus changes combined. In one of the two experimental conditions, the surprise event was kept short (affecting Steps 18–15 of the countdown = 800 ms) and configurally simple. In the other condition, the stimulus changes were more extended (affecting several steps of the countdown beginning with the first and ending with the last = 4 s) than those in the short/simple condition and in the previous experiments (on average, 1.6 s in Experiment 1 and 2.5 s in Experiment 2), and they deviated in a complex, unpredictable way from the audiovisual pattern presented during the baseline. We reasoned that this complex deviation might be less readily accommodated to the previously established schema.

After each surprise trial, the participants were presented with 11 index cards asking for the occurrence of surprise and surprise-related processes postulated in the surprise model described in the introduction (e.g., interference, attention capture, and forgetting of the letters caused by the surprising event). In addition, as in Experiment 2, the participants were again asked to graph the course of their surprise feeling across time. Given the dearth of surprise displays observed in Experiments 1–3, we were curious what the participants themselves thought about their facial reactions to the surprise event. Therefore, after the first two questions (asking for experienced surprise and interference), they were also asked the following: “Do you think your surprise showed on your face? (yes/no). If yes, how did it show? (My eyebrows went up; my eyes widened; my jaw dropped/my mouth went open; I blinked; other)”.

Results and Discussion

Surprise phenomenology and memory performance. Results concerning the experience of the surprise event replicated and extended those obtained in the previous experiments. The mean surprise rating in the first surprise trial was 53; all participants were at least minimally surprised (rating > 0), and 68% had ratings > 50. A 2 (experimental condition: long/complex vs. short/simple) $\times$ 5 (repetition of the surprise event) analysis of variance (ANOVA) with repeated measures on the second factor revealed that surprise intensity declined significantly across the critical trials from $M = 53$ in the first trial to $M = 13$ in the third trial, after which it remained at this level, $F(4, 80) = 26.1$, $p < .001$, Huynh–Feldt $\varepsilon = .81$. In contrast, experimental condition (short/simple vs. long/complex surprise event) had no significant effect, and the interaction was also nonsignificant (Fs < 1).

Parallel results were obtained in the first surprise trial for the maxima of the individual surprise curves ($M = 67$), as well as for the ratings of experienced interference ($M = 47$), distraction (50), confusion (43), attention capture (50), and forgetting (52). In all cases, the ratings declined significantly with the repetition of the stimulus changes (all repetition effects were significant at $p < .01$ or better) and reached a low bottom level at around the third repetition. The effects of experimental condition and the interaction effects were nonsignificant. In contrast to expectations, there was also no significant effect of experimental condition on the estimated duration of surprise ($M = 1.14$ s according to the direct rating), although this measure, too, declined significantly with the repetition of the surprise event, $F(4, 80) = 7.3$, $p < .01$. Finally, startle [German: “erschreckt”] was rated as low ($M = 28$ in the first critical trial).

The effectiveness of the surprise induction was additionally confirmed by the analysis of the expectedness ratings (“Did you expect a change of stimulus presentation in this trial?”). In the first critical trial, the participants clearly had not expected the surprise event to happen ($M = 10$), whereas in the second and in the following critical trials, the mean expectancy ratings ranged between 60 and 70. Finally, after the first occurrence of the surprise event, 36% of the participants said that they wondered about what had happened, and 70% said that they spontaneously inferred that the event was a part of the experiment.

The participants’ reports about forgetting, interference, attention capture, and distraction were partly supported by their performance at the memory task. In the first surprise trial, memory performance decreased from an average of 66% correctly recalled letters during a baseline period (the 10 trials prior to the first surprise trial and all trials between the critical trials) to 57% correct, $t(21) = 2.0$, $p < .05$ (one-tailed). Fifteen (68%) of the participants showed a performance decline. In the second and the subsequent critical trials, the performance decline was no longer significant.

Facial displays. Two (9%) of the 22 participants showed a component of the surprise display (one eyebrow raising, one eye widening) in the first surprise trial, and two more showed eyebrow raising to, respectively, the second and fourth repetition of the surprise event. Of the additional coding categories, only smiling/laughter was observed, which occurred in 11 (50%) of the participants in the first and twice each in the second, third, and fourth critical trial. This may reflect that the participants found the surprise event more amusing than in the previous experiments, that the level of sociality was higher, or both (cf. Hess et al., 1995).

Beliefs about expression. In striking contrast to the objective codings, 77% of the participants in the first surprise trial believed that they had shown one or more components of the facial surprise display (59% eyebrow raising, 55% eye widening, 14% mouth opening). In addition, 36% thought that they had blinked, and 5% reported that they had shown other movements in response to the surprising event, although this, too, was not confirmed by the video codings. In the second to fifth critical trial, the percentage of participants who believed that they had shown at least one component of the surprise display (of those who still felt surprised) was 46%, 32%, 45%, and again 32%, respectively. The effect of repetition was significant, $F(4, 80) = 4.73$, $p < .01$, Huynh-Feldt $\varepsilon = 1.0$, whereas the effect of experimental condition (short/simple vs. long/complex) was not, $F(1, 20) = 1.39$, $p = .25$, as was the interaction ($F < 1$).

Conclusions

Experiment 4 added to the previous experiments in three ways. First, the insufficient-duration hypothesis was not supported: There was no evidence for a greater probability of facial displays of surprise in response to a longer and more complex unexpected event. Second, we obtained additional support for the presence of surprise in most participants: Particularly in the first surprise trial, the participants reported considerable interference, distraction, confusion, attention capture, and forgetting of task material induced by the surprising event. This finding was verified by an objective decline of performance on the memory task. In addition, the participants rated the surprising event in the first surprise trial as highly unexpected, and they reported the occurrence of investigative processes and causal attributions. Third, we found that despite the low incidence of visible facial surprise expression, the majority of the participants believed that they had shown at least one component of the surprise display. We delay discussion of this—surprising—finding to Experiment 5.

Experiment 5: Embedding Surprise Displays Into Orienting Movements

Today, the most widely assumed evolutionary function of the facial displays associated with emotions is communication to conspecifics (e.g., Ekman, 1997; cf. Fridlund, 1994). In contrast, Darwin (1872/1998) proposed that emotional displays evolved primarily because of their nonsocial functions. With respect to surprise, Darwin suggested that eye widening and eyebrow raising evolved primarily to aid the rapid localization and visual investigation of an unexpected event. Similar proposals have been made by others (e.g., Andrew, 1963; Fridlund, 1994; see also the reviews and discussions in Ekman, 1979; Smith & Scott, 1997). On the basis of these considerations, it appears possible that facial displays of surprise occur preferably if the localization and direct investigation of the surprise-eliciting event require a visual search including rapid reorientation of the eyes, head, or body toward the event. In support of this hypothesis, Blurton Jones and Konner (1971) found that brow raises in children in response to a clock buzzer that suddenly went off during story time were more frequent if the clock was hidden behind some object than when it was clearly in view.

This hypothesis, which can be regarded as another modification of APT (but see the General Discussion), was reexamined for adults in Experiment 5. In a manner similar to that in Experiment 4, the participants worked on a memory task in which they were surprised by an unannounced sequence of tones. However, in contrast to Experiment 4, the tones were played through a loudspeaker located to the right and above the eye level of the participants. As a consequence, eye and head movements to the right and slightly upward were necessary to visually explore the sound source in an optimal way. We also attempted to influence (facilitate vs. inhibit) this visual exploration tendency experimentally (see *Method*).

An additional goal of Experiment 5 was to verify the findings of Experiment 4 concerning participants' beliefs about their surprise displays.

Method

Participants. Participants were 20 students (13 female, 7 male) with a mean age of 23.7 years from the same pool as in the previous experiments. They were randomly assigned to the two experimental conditions (facilitation vs. inhibition of the visual exploration tendency).

Procedure. A short-term memory paradigm similar to that in Experiment 4 was used. The memory task comprised 20 trials, with the surprise event occurring in the last trial. In contrast to Experiment 5, the surprise event was exclusively auditory in nature: It consisted of an irregular sequence of 20 high and low tones, which were played during the 20-step rehearsal countdown through a small loudspeaker located approximately 60° to the right of the participant and 30° above eye level. The speaker was partly hidden behind a wall poster to make the detection and exploration of the sound source more difficult. Also in contrast to Experiment 4, no tones were played during the baseline trials so that the location of the speaker would not be revealed prematurely.

The experimental manipulation of the visual exploration tendency was based on the assumption that the surprising event would elicit two different investigatory tendencies: a nonsocial one (visual exploration of the sound source) and a social one (asking the experimenter about the significance of the event). In the facilitation condition, the experimenter sat at a table to the right of the participant and, hence, in a direction congruent with the location of the hidden speaker. In the inhibition condition, the experimenter sat to the left of the participant. As a consequence, in the facilitation condition, the movements suggested by the nonsocial exploration tendency were compatible with those suggested by the social one, in that both involved turning to the right. In contrast, in the inhibition condition, the two exploratory tendencies suggested incompatible movements.

Immediately after the surprise trial, the participants were presented with three index cards asking for the intensity of felt surprise, perceived facial changes, and the duration of felt surprise. Behaviors were coded as before, but two new categories were added: eye or head movements toward the loudspeaker and turning to or asking the experimenter about the significance of the event.

Results and Discussion

With the exception of estimated surprise duration, there were no significant differences between the two experimental conditions; therefore, experimental condition is ignored.

Experience of surprise; interference with the parallel task. Mean intensity of felt surprise was 61.3 ($SD = 21.6$, $Mdn = 61$), with all but one participant checking numbers > 0 and all but two checking numbers ≥ 50 on the 100-point scale. Estimated duration of surprise was 1.8 s ($SD = 1.1$). Memory performance decreased from an average of 79% correctly recalled letters during baseline (the 10 trials before the critical one) to 59% correct in the surprise trial, $t(19) = 3.74$, $p < .001$, with 15 (75%) of the 20 participants showing a performance decrement.

Investigative activities. In line with expectations, the majority of the participants (12, or 60%) showed at least one of the two hypothesized investigative activities. Seven participants looked to the source of the surprise event. Eleven turned to and looked at the experimenter; nine of these asked the experimenter about the significance of the tones. Six participants showed both behaviors.

Facial displays. We observed one instance each of eyebrow raising and mouth opening and three cases of eye widening, each shown by a different participant; hence, 5 participants (25%) showed a component of the surprise display. This percentage is significantly higher (Fisher exact probability test, $p < .05$) than that observed in Experiments 1–4 combined (8 of 124 participants, or 6.3%), although only marginally higher ($p < .09$, one-tailed) than in Experiment 4 (9%). In addition, the surprise displays were preferentially shown by participants who looked to the loudspeaker: 4 of these 7 participants showed a facial component of surprise, as compared with 1 of 13 who did not look to the speaker; Fisher exact probability test, $p < .05$. Smiling or laughter (mostly the former) was observed in most participants (75%). All participants who showed a surprise expression also smiled, but in each case the smiling occurred only 2-3 s after the surprise display. Therefore, it is unlikely that smiling interfered with the surprise display in the other participants.

Beliefs about expression. Closely replicating the findings of Experiment 4, 16 (80%) of the participants said they believed that their surprise had shown on the face in one or more of the following forms: eyebrow raising (60%), eye widening (45%), jaw drop or mouth opening (30%). To aid the interpretation of the self-reports, additionally we asked the last 17 participants whether other people would have noticed their facial expressions if they had closely watched. Thirteen of these participants believed that they had shown a surprise display; 12 of them said that others could have noticed.

Conclusions

Experiment 5 yielded two main findings. First, in line with the visual exploration hypothesis, the frequency of surprise displays increased significantly compared to the previous studies if the direct exploration of the surprising event required visual search. This result replicates the findings of Blurton Jones and Konner (1971) for adults. However, the incidence of surprise displays was still small (25%) and only one-component displays were observed. In evaluating these results, it must be considered that the attempt to instigate visual search was only partly successful. When visual exploration occurred, surprise displays were more frequent (57%).

Second, Experiment 5 replicated and extended the results from Experiment 4 of a dissociation between displays of surprise and participants' beliefs about their displays. Theoretically, this finding can mean two things. First, it could mean that most participants reacted with minute, invisible surprise expressions to the stimulus changes (cf. Tassinary & Cacioppo, 1992), and people are sensitive observers of even such invisible expressions. Second, it could mean that the participants' self-reports were based on a different source of information. What comes to mind here are, in particular, generalized beliefs about the association between surprise and expression (cf. Ekman et al., 1987). The finding of Experiment 5, that participants believed that their surprise expressions were visible to others, supports the second explanation. At least, participants believed that their facial expressions were more intense than they in fact were.

Experiment 6: Generalization to a Different Surprise Event

The main aim of Experiment 6 was to test whether the findings of the previous studies are restricted to surprising events of the kind staged in these studies (simple audiovisual changes) or can be generalized to other kinds of surprising events. In choosing a generalization event, we wanted in particular to meet any remaining concerns that the surprise induced in the previous studies, including Experiment 3, was still not *intense* enough to elicit a facial display. To address this concern, we staged a surprise situation that seemed intuitively powerful but still permitted strict experimental control: Participants were secretly photographed while they rated a series of pictures of faces on the monitor, and their own picture was presented to them as the last in the series. In a pretest, where we described this event to 33 participants and asked them to estimate their likely reactions, it received not only high surprise ratings, but most participants also believed that their surprise would show strongly on their face (rating of $M = 70$ on a 0–100-point scale). The pretest also suggested that the described surprising event would be experienced as highly amusing, and thus as a pleasant surprise. Furthermore, confrontation with one's own face is held by some authors to be a powerful social stimulus (e.g., Wicklund & Frey, 1980), that should therefore have higher personal relevance than the surprise events staged in the preceding studies.

The second aim of Experiment 6 was to retest the display-rules hypothesis (cf. Experiment 3), again by varying the level of sociality. We used this opportunity to simultaneously test yet another possible modification of APT: that the surprise display occurs only when the level of sociality is *high* (more detail is given in the *Method* section).

Finally, to further clarify the findings of Experiments 4 and 5 concerning beliefs about facial expression, we asked the participants of Experiment 6 to rate the intensity with which surprise had shown on their face and to describe the facial changes in their own words.

Method

Participants. Participants were again 23 students (13 female, 10 male) at the University of Bielefeld, whose mean age was 25.3 years. They were randomly assigned to one of two experimental conditions, with 11 in the nonsocial and 12 in the social condition. Seven additional participants had to be excluded from the data analyses, 3 because they did not recognize themselves on the monitor and 4 because of equipment problems.

Procedure. Participants were told that the goal of the experiment was to test whether photographs of faces are judged differently when presented in different media. In the first phase of the experiment, they judged 16 pictures of faces presented on the monitor. To ensure that the participants paid attention to the identity of the depicted person, they first indicated via a keypress whether or not the face appeared familiar to them. Subsequently, they rated the person on two trait scales (conscientious and well balanced). The photographs were black-and-white and color pictures of faces showing mostly a neutral expression and included a few pictures of well-known politicians and actors. A spy camera hidden in a book case next to the monitor transmitted an image of the participant's face to the adjoining room. There, a confederate made a photograph of the participant's face from the incoming video stream with the help of video capturing software and edited the picture to make it similar in appearance to the other photographs.

For the second phase of the experiment, the participants were seated at a different table, where they judged 16 printed photographs on the same scales. This allowed the experimenter to unobtrusively download the participant's picture to the experimental computer. In the third phase of the experiment, the participants judged another series of faces on the monitor. The seventh and last face of this series was their own. Subsequently, they completed a set of rating scales and questions. In addition to items asking for surprise and surprise-related processes, these included 14 emotion or mood items (e.g., happy, angry, startled, nervous, embarrassed, tired, wakeful).

Sociality was manipulated as follows: In the alone condition, the experimenter left the room after the instruction and returned only when the participant, after the first and second series of photographs, pressed a signal button. The experimenter then briefly explained the next phase and left again. After the surprise trial, he waited for 30 seconds before he returned to the room. In the social condition, the experimenter remained in the room throughout the experiment. However, different from the manipulation of sociality used in Experiment 3, he sat to the side of the monitor table facing the participant, and noted down his or her picture ratings, that had to be made verbally in this condition. This established a continuous face-to-face interaction between experimenter and participant.

Results and Discussion

Evidence for surprise. Sociality had no significant effect on most variables and will therefore only be mentioned for the exceptional cases. Most participants judged the appearance of their own face on the monitor as very surprising, with all 23 scoring $\geq$30 on the 0 (*not at all surprised*) to 100 (*extremely surprised*) scale, and 16 $\geq$ 60; $M = 69$ ($SD = 21$, $Mdn = 70$). Similar results were obtained on scales asking for astonishment ($M = 67$) and amazement ($M = 66$). To obtain more information on the meaning of the surprise ratings, the participants were also asked to recall a "highly surprising" event from their past and to compare it with the experimental event. On average, participants judged the intensity of surprise caused by the experimental event to be 70% of that of the recalled event ($SD = 38$); 10 gave percentage scores $\geq$70% and 5 > 100%. The effectiveness of the surprise induction was further supported by the ratings of attention capture ($M = 72$) and confusion (55), and by reports about spontaneous explanatory search (87% of the participants). Also, the participants typically had not expected that anything unusual would happen during the experiment, $M = 31$. The average estimated duration of surprise was 4 s (with 8 responses in the ">5-s" category fixed at 6 s).

The surprising effect of the appearance of one's own face was also reflected behaviorally: (a) The response to the "familiar face" question was significantly retarded relative to baseline (the average RT of the 10 pictures preceding the critical item), $t(22) = 4.3$, $p < .001$. Eighteen participants showed an RT increase from baseline; of those who did not, 2 recognized themselves only during the first trait judgment. (b) Ten of the 12 participants in the social and 4 in the nonsocial condition made spontaneous verbal exclamations suggestive of surprise, such as "hey" or "that's me," $p < .05$ (Fisher exact probability test). (c) Seventy-four percent of the participants showed evidence of either a visual search (e.g., taking a second look at the picture) or, more typically, a verbal search (asking the experimenter).

Facial expression of surprise and beliefs about expression. We coded the first 10 s after the onset of the surprise event. We observed one case of eyebrow raising and one case of eyebrow raising plus mouth opening, one in the social and the other in the nonsocial condition (8.7%). Nonetheless, all participants believed that they had shown a surprise expression. In this study, we first asked them to state how strongly their surprise had shown on the face on a scale ranging from 0 (*did not show at all*) to 100 (*showed extremely strongly*). The mean rating on this scale was $M = 78$ ($SD = 18$); 19 (82%) of the participants had scores $\geq$70. Second, we asked the participants to describe in their own words how surprise had shown on the face (in a way visible to others). The most frequently named expression was smiling/laughter (34%), followed by "wide eyes" (30%; this description may have been meant to include raised eyebrows), brow raising (26%), and mouth opening/jaw drop (9%). At least one of the three classical surprise components was named by 52%.

Other affects and nonverbal behaviors. As suggested by the pretest, the other strong emotion elicited by the sudden appearance of one's own face was amusement, $M = 75$ ($SD = 19$). This was presumably also reflected in a high happiness rating on the mood questionnaire ($M = 72$). With the exception of a set of items concerned with wakefulness and relaxation, the means for all other emotion and moods items were low (e.g., angry = 4, embarrassed = 27, and startled = 30 on the 100-point scale).

The most frequently observed facial expression was smiling/laughter, which occurred in 22 of the 23 participants. However, as in Experiment 5, the two participants who showed a partial surprise display smiled only several seconds after this display. The intensity of the mirth expression was higher in the social condition (11 cases of laughter) than in the alone condition (5 times). If smiling is coded as 1 and laughter as 2, the difference between the means of the two conditions (1.9 vs. 1.3) is significant, $t(21) = 2.6$, $p < .05$. Because the occurrence of amusement was independently ascertained in this study, this finding suggests that the presence of the experimenter increased the tendency to express amusement (cf. Hess et al., 1995).

Conclusion

Experiment 6 tested whether the previous findings could be generalized to a different, intuitively powerful surprise situation. Confirming this intuition, the unexpected appearance of their own face on the monitor was, on average, rated as having 70% of the intensity of a recalled, highly surprising event. Nonetheless, the results concerning facial expression replicated all of the central findings of the previous experiments. First, only 2 of the 23 participants (9%) showed a surprise display, and in both cases it was partial only. This finding refutes once again the insufficient-intensity hypothesis (Experiment 3). Second, the manipulation of sociality had no effect on the surprise display. This finding speaks once more against the hypothesis that surprise displays were inhibited or masked (Experiment 3); a conclusion that receives additional support from the finding that sociality *did* affect spontaneous verbal exclamations suggestive of surprise, as well as mirth reactions. Furthermore, the finding that even a face-to-face interaction failed to bring forth more of a surprise display speaks against yet another hypothesis: that too low a level of sociality prevented the surprise expression from occurring in the preceding studies. Third, the finding that the estimated duration of surprise (on average, 4 s) was much longer than in Experiments 4 and 5 (1.1 s and 1.8 s, respectively) speaks once more against the insufficient-duration hypothesis (cf. Experiment 4). Finally, as in Experiments 4 and 5, the participants typically believed that they had shown components of the surprise display. The frequency of spontaneously mentioned surprise components was about 30% less than that obtained by the checklist method in Experiments 4 and 5, but they were still listed by the majority. Furthermore, most participants believed that their surprise expression had been intense. This speaks further against the hypothesis that the self-reports about facial expressions were based on invisible, minute facial changes.

Finally, we found that, apart from amusement, other strong emotions were not elicited by the surprising event staged in Experiment 6. However, because the amusement ratings were as high as those of surprise, proponents of APT could at this point raise the objection that the expression of surprise did not occur with higher frequency because the feeling of amusement, or the facial display occasioned thereby, overruled surprise or the associated expression. Regardless of the merits of this explanation in other cases, we do not regard it as convincing in the present case for both theoretical and empirical reasons. That is, APT does not predict that just *any* facial movement or feeling that co-occurs with surprise interferes with the facial display of surprise; only incompatible movements and strong incompatible emotions do. Of the facial components of surprise, at least eyebrow raising is, however, not incompatible with smiling. Also, the feeling of surprise is not incompatible with that of amusement; on the contrary, surprise is often regarded as a precondition or a component of amusement (e.g., Suls, 1971; see also Deckers, 1993). For this case—the co-occurrence of two compatible emotions—APT predicts that signals for *both* facial displays are sent to the face, resulting in a facial blend (e.g., raised eyebrows in a smiling face; Ekman, 1972). However, this was not observed. Furthermore, in both Experiments 5 and 6, the surprise displays (of the few participants who showed one) preceded smiling by at least a second, suggesting that the feeling of surprise was present in pure form at least briefly before amusement set in—long enough, we suggest, to manifest itself on the face.

Experiment 7: Testing for Invisible Brow Raisings

Strictly speaking, the results of Experiments 1–6 pertain only to surprise displays that are visible to observers. Although it may seem implausible, given the high intensity of surprise induced in Experiments 3 and 6, it is still conceivable that many participants showed minute surprise expressions that were below the coders' threshold of awareness (Tassinary & Cacioppo, 1992). We conducted Experiments 7 and 8 to examine this possibility by measuring facial EMG. EMG recordings are able to detect muscle movements that are invisible to the naked eye (Fridlund & Cacioppo, 1986). Experiment 7 was a replication of Experiment 6 (without the sociality manipulation), whereas in Experiment 8 we used a surprise paradigm comparable with those used in Experiments 1–6.

Method

Participants. The final sample of Experiment 7 consisted of 28 students (13 female, 15 male) with a mean age of 22.4 years of various disciplines—mostly nonpsychology—at the University of Greifswald. Eight additional participants were excluded from the data analyses, four because they did not recognize their face on the monitor, three because of problems with the EMG measurement, and one because of a procedural error.

Procedure. The experiment was conducted by two experimenters, one male and one female. The procedure was similar to that of Experiment 6. To distract the participants from their facial muscles, we told them that we were interested in subtle changes of blood flow in the face during picture viewing. Although subjectively unnoticeable, these physiological reactions could presumably be detected by temperature sensors that would be placed on selected places of the face. After the participants were seated in front of the computer monitor, a snapshot of their face was secretly taken from the incoming video stream of the spy camera. One experimenter (always the same sex as the participant) then attached the EMG electrodes, while the other experimenter, who was separated from the participant by a room partition, edited the picture and transferred it to the experimental computer through a parallel link. After a relaxation period, 38 black-and-white photographs were presented. The participant's task was to view the photographs and to indicate whether the depicted person appeared familiar. Pictures were separated by an intertrial interval of 3.3 s, during which a blank screen was shown. Level of sociality was kept constant at a low level, as it had no effect in Experiment 6 (nor in Experiment 3). That is, the same-sex experimenter stayed in the room behind the partition and busied himself or herself with supervising the EMG apparatus. The picture of the participant's face was shown in Trial 38 and remained on the screen for 10 s. Subsequently, the experimenter moved to the participant's table and asked the postexperimental questions, which were largely the same as those used in Experiment 6: The mood scale was omitted, the free listing of perceived facial surprise components was again replaced by checking components on a list, and a question asking for the time when the face had been recognized was included. After the experiment, the participants were debriefed about the true nature of the physiological recordings.

EMG measurement. Of the muscles involved in the surprise display, we only considered the frontalis muscle, responsible for eyebrow raising. The muscle responsible for eye-widening (musculus levator palpebrae superioris) retracts over the eyeball into the orbit; therefore its activity cannot be measured with surface EMG (although it can be measured with needle electrodes; cf. Aramideh, Ongboer de Visser, Devriese, Bour, & Speelman, 1994). We also neglected mouth opening/jaw drop because there are no established guidelines for its EMG measurement and because a pretest (using voluntary jaw dropping) in which we attempted to index this facial movement by the relaxation of the masseter muscle was unsuccessful. However, given that eyebrow raising was the most frequent visible facial surprise component observed in Experiments 1–6 and in the previous study by Reisenzein (2000a), the frontalis muscle seemed the most promising place where to look for invisible surprise displays. In addition, we recorded EMG activity over corrugator supercilii (responsible for brow knitting), partly to control for possible artifacts in the frontalis EMG that may have been due to corrugator movements, and we recorded EMG activity over zygomaticus major to detect possible invisible smiles.

We recorded the EMG signals using the Vitaport II recorder (Temec Instruments B.V., the Netherlands). Miniature (0.3-cm) bipolar Ag/AgCl electrodes were placed on the left side of the face in accordance with the guidelines of Fridlund and Cacioppo (1986). Amplifiers were set at a theoretical resolution of 0.23 μV. We filtered the EMG signals with an 8Hz high-pass and a 400Hz low-pass hardware filter and digitized them at 1024 samples per second. Offline, we filtered the recorded EMG signals again with a 16Hz high-pass filter to attenuate blink and eye movement artifact (van Boxtel, 2001), as well as with a 50Hz notch filter to eliminate possible power line interference. Subsequently, the EMG signals were full-wave rectified and then smoothed with a flat 10Hz low-pass filter. For each channel, we computed the mean EMG amplitudes for the 20 consecutive 0.5-s intervals following stimulus onset in the surprise trial. To estimate the baseline variability of EMG activity, we computed the standard deviation of the means of the 11 baseline (1-s prestimulus) periods consisting of the surprise trial and the 10 preceding trials. Using this standard deviation estimate and the mean of the 1-s baseline immediately preceding the critical trial, we then individually standardized the poststimulus EMG means (cf. Hess & Blairy, 2001).

Results

Evidence for surprise. The subjective data replicated those of Experiment 6. Mean rated surprise on the 0–100-point scale was 79 (SD = 14; Mdn = 80), and relative surprise in comparison to an "extremely surprising" remembered experience was 68% (SD = 63, Mdn = 50). Participants had not expected anything unusual to happen during the experiment (M = 24), were astonished (M = 68) and confused (M = 74) by the appearance of their own face, found their attention strongly captured by it (M = 74) and typically searched for an explanation (61%). Again, participants were also strongly amused (M = 82) by the surprising event. Again, the surprise stimulus caused a significant RT increase of the familiarity judgment relative to the baseline in the preceding 10 trials, $t(27) = 1.75$, $p < .05$ (one-tailed); and again, most participants eventually addressed the experimenter to ask for an explanation. Finally, according to the retrospective reports, the identity of the face was detected, on average, 1.3 s after picture onset (SD = 1.3, Mdn = .75).

Video codings and beliefs about expression. The videos of the 10-s period of the surprise trial were digitized and coded for facial expressions with a software media player. We observed four brow raises and two eye widenings; at least one of these displays was shown by 5 participants (18%). This is not significantly different from the frequency of surprise expressions obtained in Experiment 6 (9%), Fisher exact probability test, p = .30. Also similar to Experiment 6, we observed smiling/laughing in most participants (86%); with the exception of two participants who broke into a laughing fit, it consisted of smiling only, similar to the nonsocial condition of Experiment 6. Again similar to Experiment 6, the participants believed that their surprise had shown strongly on the face (M = 86, SD = 15, Mdn = 85). Similar to Experiments 4 and 5 in which a checklist method had been used, 64% believed that they had shown brow raising, 71% eye widening, and 57% mouth opening/jaw drop; 93% checked at least one surprise component. The correlation between observed and perceived expression components was close to zero; $r = -.12$ (brow raising) and .12 (at least one component shown).

EMG activity. Before the data analysis, we scanned the EMG records for movement artifacts using the video recordings and onscreen displays of the EMG. Most movement artifacts occurred in the second half of the 10-s picture presentation period and were due to the fact that about one third of the participants turned to and addressed the experimenter (which involved eye, head, and body movements and talking) before the end of the observation period. For these participants, only the first 4–9 s of the 10-s period could be evaluated. Two other participants, as mentioned, broke into a laughing fit; for these, we had to discard the remainder of the trial. One participant briefly looked to the ceiling during the later part of the surprise trial and another lowered his head and peered at the picture "from below lowered eyes," both of which resulted in an increase of frontalis EMG activity; these periods (about 1–2 s) were also discarded.

Because the main aim of the EMG measurement was to detect possible invisible frontalis activity, the statistical analysis centered on the 24 participants who did not show visible brow raising (as expected, the visible brow raisings were reflected in highly significant increases of the frontalis EMG). For these participants, on average, 8 s of artifact-free EMG were available; 11 had complete protocols for the whole 10-s period, and 19 had complete protocols for the first 5 s. The mean change of frontalis and corrugator EMG activity during the 10 s of picture presentation is shown in Figure 2. We obtained nearly

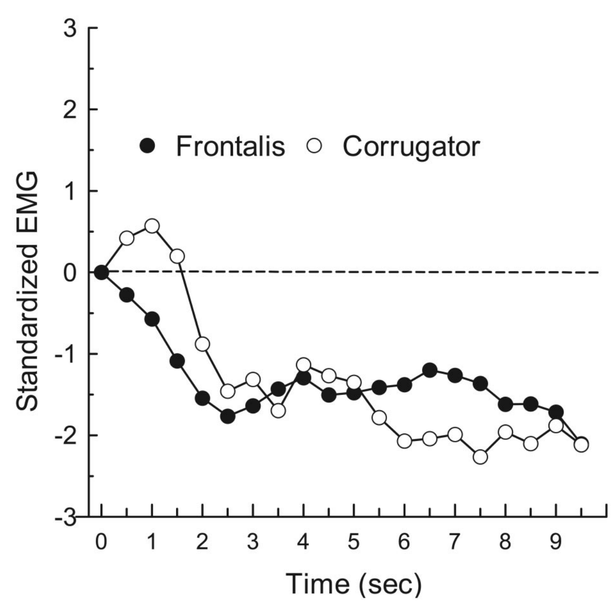

Figure 2. Average frontalis and corrugator electromyographic (EMG) responses during the surprise trial, Experiment 7.

identical findings when we only considered the 11 participants with complete protocols. As can be seen, overall there was a decrease of frontalis activity across time, as well as a decrease of corrugator activity after a slight elevation after stimulus onset. Dependent t tests comparing the (unstandardized) mean in each poststimulus interval with the prestimulus baseline mean revealed significant decreases of frontalis EMG ($ps < .01$, dfs between 23 and 12) between 1.5 s and 8.5 s and again from 9–9.5 s. The corrugator EMG also showed a significant decline from 3 s to 8.5 s; the initial increase at 0.5 and 1 s (see Figure 2) was marginally significant ($ps = .05$ and .08, respectively).

Because the mean data could have masked significant frontalis EMG increases in individual participants, we next examined the individual changes of EMG activity across time. Z values > 2 were scored as significant increases from baseline, and z values < -2 were scored as significant decreases. Using this criterion, the individual graphs could be classified into three nonoverlapping groups: Fifteen (54%) participants showed a significant decrease of frontalis EMG during at least one of the poststimulus periods but no significant increase; 7 (25%) had neither a significant increase nor a significant decrease, and 2 participants showed a significant increase but no significant decrease. Thus, the typical temporal pattern of the frontalis EMG change was either a decrease or no change. Finally, the Spearman rank correlation (used to account for possible nonlinearities) between the standardized frontalis EMG and self-rated surprise was close to zero for all 20 measurement intervals.

A significant increase of zygomaticus activity ($> 2z$) was detected in 24 of the 28 participants, the same who also showed a visible smile. With one exception, the frontalis EMG increase preceded the zygomaticus response by 1–3 s, confirming the impression gleaned from the videos and suggesting that amusement set in only a few seconds after surprise. Consistent with this interpretation, the rank correlation between the standardized zygomaticus EMG and self-rated amusement became significant ($r = .48$, $p < .01$) 4.5 s after stimulus onset, after which it remained significant or close to significant until the end of the measurement period.

Discussion

Experiment 7 largely replicated the results of Experiment 6. The unexpected appearance of their own face on the monitor was judged as strongly surprising by most participants, but only 18% showed a visible component of the surprise display. The EMG measurement suggested two additional instances of brow raising, thus raising the frequency of brow raising from 14% (video) to 21% (EMG). At the same time, however, the EMG measurement revealed a significant decrease of frontalis muscle activity in the majority of the participants with no visible surprise displays, and the correlation between frontalis EMG and self-rated surprise was essentially zero.

Although not predicted, the observed decrease of frontalis—as well as corrugator—activity in the majority of the participants is consistent with recent results reported by other authors. Stekelenburg and van Boxtel (2002) examined psychophysiological reactions to novel sounds (e.g., animal sounds, human talk, industrial and environmental noises) that were presented from time to time during a text-reading task without forewarning and thus were presumably at least somewhat surprising (subjective measures were not taken). In one of their experiments (Experiment 1), they found, exactly as we did, that the stimuli caused a decrease of frontalis activity, as well as a decrease of corrugator activity after a small initial increase. Camras et al. (2002) and Scherer et al. (2004) reported that expectancy violations in infants led to a temporary cessation of facial movements in many children. Post hoc, these findings fit well with the inhibitory effect of unexpected events on ongoing mental processes—and, consequently, the behaviors controlled by these processes—postulated in our surprise model (cf. the introduction). As noted, this model assumes that surprise-induced response inhibition serves to prepare the organism for the analysis of unexpected events (Meyer et al., 1997).

Finally, the EMG findings further clarify the interpretation of the participants' reports about perceived surprise displays. As noted, one possible explanation of these self-reports is that they were based on minute, invisible, facial changes. This hypothesis was already thrown into doubt by the finding that the participants believed their surprise displays to be visible to others (Experiments 5 and 6) and by the high perceived intensity of the expression (Experiment 6, replicated in Experiment 7). Further refuting this hypothesis, Experiment 7 revealed that there was no correlation between observed and reported brow raisings, that 92% of the facially unresponsive participants showed no significant frontalis EMG increase, and that 63% even showed a significant decrease.

Experiment 8: Once More With EMG

To test whether the findings for the frontalis EMG obtained in Experiment 7 generalize to the surprise paradigms used in Experiments 1–5, we conducted a final experiment. Also in this study, once again we varied the factors *intensity of surprise* and *sociality* to examine their possible effects on EMG activity. Participants were randomly assigned to the resulting four experimental conditions.

Method

Participants. The participants were 23 students at the University of Bielefeld. The surprise induction method was similar to that used in Experiment 1. The main difference was that the participants worked on a numerical addition task instead of a choice reaction task. In each trial, they had to add three numbers that appeared on the screen for 5.3 s. Subsequently, a solution number was presented for 3 s, and the participants decided whether it was correct. Trials were separated by 3 s, during which a blank screen was shown. In Trial 25, 2.5 s after the presentation of the numbers, a salient change of the mode of stimulus presentation occurred similar to the strong-background-change condition of Experiment 1: a repeated change of the color of the screen background and an inversion of the text color, accompanied by a sequence of tones.

Design and procedure. The experiment had a 2 (sociality: experimenter present vs. absent) × 2 (task difficulty: low [one-digit numbers] vs. high [two-digit numbers]) design. We varied task difficulty to influence the degree of interference caused by the surprising event and, thereby, the intensity of felt surprise (see Experiment 2). Observable facial reactions were coded as before. We measured the frontalis EMG with the Vitaport I, a precursor model of the recorder used in Experiment 7. This recorder features a dedicated EMG channel that integrates (rectifies and smoothes) EMG online. The integrated signals were digitized at 256 Hz. In this study, we only measured activity over the frontalis muscle. We evaluated the frontalis EMG from the onset of the surprise event in the critical trial until the end of the number presentation (3 s later) in 0.5-s intervals. The means of each measurement interval were individually standardized using the mean of the 1-s interval preceding the surprise event, and the standard deviation of the means of the 11 baseline (1-s prestimulus) periods consisting of the surprise trial and the 10 preceding trials. Before the statistical analyses, we scanned the EMG records and excluded periods with artifacts. Artifacts that were due to body, head, or eye movements were very rare in this study, but we had to exclude periods with blinks (on average, 1.5 per participant), because no high-pass filtering of the raw EMG had been used.

Results and Discussion

Because of a procedural error, surprise ratings were not collected in this study. However, Reisenzein and Studtmann (2006, Experiment 2), who used the same paradigm, found that the stimulus changes caused surprise in most participants and that higher surprise was felt in the difficult than in the easy-task condition ($M = 75$ vs. 48, $p < .001$). Furthermore, the stimulus changes caused a significant increase of RT and a significant performance reduction relative to baseline, $t\text{s}(22) > 2.73$, $p\text{s} < .05$.

Video codings revealed 5 participants (23%) who showed at least one surprise component (four eyebrow raises, one eye widening). Three other participants, instead of raising their eyebrows, frowned in response to the surprising event; these participants were excluded from the subsequent analyses. We then examined the EMG protocols of the 16 participants with no visible eyebrow movements for invisible frontalis activity, using the same procedure as in Experiment 7. This analysis suggested one additional case of eyebrow raising (standardized frontalis EMG $> 2z$) in the person who showed visible eye widening. In contrast to Experiment 7, the analysis of the EMG protocols of the remaining participants and of the average EMG indicated unchanged activity. One possible, if speculative, explanation for this difference to Experiment 7 is that the surprising stimulus changes were too brief to allow further sensory exploration that may have benefited from "facial stilling." Smiling occurred in 6 participants, 2 in the nonsocial and 4 in the social condition. There was no significant difference in the frequency of visible surprise displays between the social and the alone condition (2 vs. 3), nor between the easy- and difficult-task conditions (3 vs. 2). There were also no significant effects of sociality and task difficulty on the frontalis EMG in the six 0.5-s poststimulus intervals.

General Discussion

The first goal of the studies reported here was to provide further evidence on the relation between surprise and facial expression. As concerns this issue, the results of the studies provide evidence for several types of *dissociation*: (a) between the mental state of surprise and its traditional facial display, (b) between surprise and expression with respect to their reactivity to experimental manipulations, (c) between the different components of the surprise display, and (d) between the occurrence of surprise displays and participants' beliefs about their occurrence.

Dissociation Between Surprise and Facial Expression

This is the theoretically most important type of dissociation found. The pertinent evidence can best be summarized by referring to the theoretical model of surprise described in the introduction, on which the induction and the measurement of surprise were based (see Figure 1). Without repeating the details, self-report and behavioral data collected in the different studies suggested that all of the surprise-related processes postulated in the model—the appraisal of unexpectedness, the feeling of surprise, the interruption of processing, and so forth—occurred in the majority of our participants.[7] In addition, the presence of surprise is supported by the nature of the experimental inductions of surprise, which can claim theoretical and intuitive validity. On the basis of these data and theoretical considerations, the principle of "inference to the best explanation" (Harman, 1989) warrants the conclusion that the mental state of surprise was indeed present in most participants: No alternative hypothesis provides for a better explanation of the complete pattern of subjective and (nonfacial) behavioral data.

At the same time, visible or EMG-detected facial expressions of surprise occurred only in a small minority of our participants. Table 1 summarizes the pertinent findings. As can be seen, overall only 11% of the 220 participants showed a *visible* facial surprise expression (at least one component), with a range of 4% (Experiment 2) to 25% (Experiment 5). Results from Experiments 7 and 8 indicate that the low incidence of visible expressions cannot be plausibly attributed to invisible displays: Measurement of the frontalis EMG suggested only one to two additional invisible brow raises; and in Experiment 7, 54% of the participants even showed a significant decrease of frontalis activity—exactly the opposite of what APT predicts. The present findings therefore document an even more extreme dissociation between surprise and facial expression in adults than Reisenzein's (2000a) study.

The dissociation between surprise and facial expression was also reflected in their differential reactivity to experimental manipulations of surprise intensity (degree of schema discrepancy, Experiment 1; task difficulty, Experiments 2 and 8; repetition of the surprise event, Experiments 1 and 5). Whereas these manipulations had the predicted effects on subjective and on nonfacial behavioral measures of surprise, they had no statistically reliable effects on facial expression.

Dissociation Between the Components of the Surprise Expression

Of the 27 observed cases of visible or invisible surprise displays, 24 consisted of a single component: most frequently, eyebrow raising; the remaining 3 were two-component expressions. The "complete" surprise face was never seen (see Table 1). These results are again similar to, if more extreme than, previous findings by Reisenzein (2000a), who observed 54% single-component (mostly brow raising), 31% two-component (mostly brow raising and eye widening), and only 15% three-component displays. The findings are also in accord with data by Carroll and Russell (1997) on the surprise displays of movie actors (although these were posed rather than spontaneous expressions).

[7] Additional support for this conclusion stems from psychophysiological studies which found that unexpected events of the type staged in Experiments 1–5 and 8 also cause physiological orienting responses (skin conductance responses and heart rate changes; e.g., Maher & Furedy, 1979; Niepel, 2001; Siddle & Jordan, 1993).

Table 1
Summary of the Main Results of Experiments 1–8

Experiment	Surprise paradigm	Experimental manipulations	*n*	Surprise expression					Beliefs about expression (% of participants)			
				Brow	Eye	Jaw	≥1	≥2	Brow	Eye	Jaw	≥1
1	Audiovisual change	Surprise intensity	57[a]	2	1	0	3 (5%)	0	—	—	—	—
2	"Computer breakdown"	Surprise intensity	25	0	1	0	1 (4%)	0	—	—	—	—
3	Rule violation	Sociality	22	2	0	0	2 (9%)	0	—	—	—	—
4	Audiovisual change	Duration of event	22	1	1	0	2 (9%)	0	59	55	14	77
5	Audiovisual change	Ease of visual orienting	20	1	3	1	5 (25%)	0	60	45	30	80
6	Picture of own face	Sociality	23	1	1	1	2 (9%)	1	26[b]	30[b]	9[b]	52[b]
7	Picture of own face	—	28	4/6[c]	2	0	5/7 (18/25%)[c]	1	64	71	57	93
8	Audiovisual change	Sociality, intensity	23	4/5[c]	1	0	5/5 (22/22%)[c]	0/1[c]	—	—	—	—
1–8			220	15/18[c]	10	2	25/27 (11/12%)[c]	2/3[c]	52	50	28	76

Note. For Experiments 1 and 5, data from the first critical trial are used. Dashes indicate that data were not collected.
[a] For which video data were available. [b] Free listings of perceived expression components. [c] Video only/frontalis EMG included.

It appears that, as originally formulated, APT does not allow for incomplete emotion expressions except in the sense that components of a display are selectively inhibited or masked or are too weak to be visible. Correspondingly, in their *FACS Investigator's Guide*, Ekman et al. (2002, p. 174) did not list single-component displays among the expressions of surprise, and several previous investigators also required the presence of at least two of the three facial surprise components to code surprise (e.g., Bennett et al., 2002; Reissland et al., 2002). According to this stricter criterion, only 3 (1.3%) of our 220 participants showed a surprise display. The predominance of partial surprise displays may therefore signal the need for yet another modification of APT, for it could mean that the different components of the surprise expression are controlled by separate mental processes rather than by a unitary motor program (e.g., Ortony & Turner, 1990; Smith & Scott, 1997). However, note that an alternative explanation more in line with APT is possible: The different components of the surprise display could have different response thresholds, with eyebrow raising appearing first. Assuming that the hypothetical additional factor necessary for a surprise display (in addition to the presence of surprise and the absence of inhibition) was not present in our experiments, one would then expect this display to occur not only infrequently in these situations but also in partial form. In any case, it is interesting to note that lay people do allow for the occurrence of partial surprise displays (Experiments 4–7).

Dissociation Between Surprise Displays and Beliefs About the Displays

A fourth type of dissociation—a dissociation between surprise displays and beliefs about them—was documented in Experiments 4–7. These studies found consistently that the participants grossly overestimated their surprise expressions: In contrast to the results of the video codings and the analyses of the EMG data (Experiment 7), most participants believed (a) that their surprise had strongly shown on the face (mean intensity ratings ≈ 80 on the 0–100-point scale; Experiments 6–7)—in any case, in a way visible to others (Experiments 5 and 6)—and (b) that the surprise expression included one or more of the traditionally posited features (i.e., eyebrow raising, eye widening, mouth opening/jaw drop).

Converging evidence from Experiments 4–7 indicates that the reports about perceived surprise expressions were not based on visible or invisible facial displays (see the discussion of Experiment 7). Therefore, the participants must have relied on a different source of information. As we hinted earlier, we believe that the participants based their expression reports on schemas or generalized beliefs about the emotion–face association (see also Rimé, Phillipot, & Cisamolo, 1990). More precisely, we propose that they *inferred* their probable facial expression from their feelings of surprise (minor premise) and from generalized beliefs about the facial expression associated with surprise (major premise): They reasoned that, because they felt surprised, and because surprise is associated with a characteristic facial display, they must have shown this display. Experimental support for this hypothesis was obtained by Reisenzein and Studtmann (2006), who found that an experimental manipulation of surprise intensity, although not influencing expression, significantly affected participants' beliefs about their facial expression. Apparently, then, people are relatively insensitive to their own facial displays of surprise (or at least their absence) but highly susceptible to schema-based beliefs about the emotion–face association.

This conclusion, if correct, may help to explain the discrepancy between both folk–psychological and scientific beliefs about the relation between surprise and facial expression (cf. the introduction), and the present findings. As already suggested in the introduction, the cognitive representation or schema of surprise to which people recur when they judge the association between surprise and facial expression (e.g., Ekman et al., 1987), does not seem to reflect the statistically modal but the ideal case of surprise, where the surprise syndrome is present in full-fledged form (Horstmann, 2002). The question then is how this ideal-type schema is acquired in the first place and why it is not corrected by experience. The present findings suggest a partial answer to these questions: If people are not sensitive to their own facial displays (e.g., simply because they usually do not attend to them), then a central source of information available for acquiring veridical beliefs about the emotion–face association is neglected. In particular, people will then miss cases where surprise is present but facial expression is not. Worse, reliance on the ideal-type schema even in personal experiences of surprise leads into a self-reinforcing cycle: Prospectively, one expects surprise displays to occur even in the statistically modal cases of surprise; retrospectively, one surmises them to have been present in these cases (see also Schützwohl & Krefting, 2001), thereby apparently confirming one's expectations.

Explanations of the Emotion–Face Dissociation

The second goal of the studies reported here was to explore possible explanations for the dissociation between the emotion of surprise and its facial display.

First, we considered method problems related to the induction and measurement of surprise. On the basis of Experiments 1 and 2, we concluded that the method artifact hypothesis can be ruled out.

Next, we examined the two substantive explanations for the observed dissociation available to APT (as originally stated): inhibition or masking of facial displays due to display rules, and insufficient surprise intensity. In our view, the display-rules hypothesis cannot explain the results, because the incidence of surprise displays was no higher in nonsocial than in social situations (Experiments 3, 6, and 8). The insufficient-intensity hypothesis, too, was not supported. First, the surprise display was insensitive to manipulations of surprise intensity (Studies 1, 2, and 8). Second, even high induced surprise (Studies 3, 6, and 7) did not result in much more of a surprise display. Third, additional analyses revealed that the frequency of surprise expressions in highly surprised participants—those with ratings ≥ 7 on the 0–10-point surprise scale or ≥ 70 on the 0–100-point scale—was nearly identical (12.6%) to that of surprise expressions in the total sample (11%). Fourth, EMG measurements (Experiments 7 and 8) detected only very little invisible facial activity related to surprise, and in one study (Experiment 7) they even revealed a decrease of frontalis muscle activity in the majority of the participants.

Finally, we turned to modifications of APT, obtained from the original theory by adding the assumption that some other factor X, apart from surprise and the absence of deliberate control, is needed for the facial surprise display to occur. Three such modifications were examined: the insufficient-duration hypothesis, the visual-orienting hypothesis, and the insufficient-sociality hypothesis. The insufficient-duration hypothesis was not supported: An experimental manipulation of the duration of unexpected stimulus changes did not affect facial expression (Experiment 5) and even comparatively long-lasting surprise (Experiment 6) did not produce surprise expressions. The insufficient-sociality hypothesis was also unsupported (Experiment 6): Although a high level of sociality increased spontaneous verbal exclamations suggestive of surprise as well as mirth reactions, it did not increase the frequency of surprise expressions. Only the visual-orienting hypothesis found some support (Experiment 5). Admittedly this support was weak; however, the findings agree with previous results of Blurton Jones and Konner (1971). In addition, the higher frequency of (partial) surprise displays observed in some of the surprise situations staged in previous studies (cf. the introduction) may also be explainable by the visual-orienting hypothesis. On the other hand, visual orienting does not seem to be generally necessary for surprise displays (e.g., Reisenzein, 2000a). To reconcile these findings with the visual-orienting hypothesis, one would need to add something like Darwin's (1872/1998) assumption that, due to "the force of association" (p. 281), surprise displays are eventually shown even to events that do not require visual search.

Finally, it needs to be pointed out that, whatever the merits of the visual-orienting hypothesis may be, as an adjunct to or a modification of APT, this hypothesis is problematic. To fit the proposed schema of an APT modification (i.e., given absence of inhibition: surprise + factor X $\rightarrow$ facial display), the visual-orienting hypothesis must be read as follows: surprise + need for visual orienting $\rightarrow$ facial display. However, the more natural explication of this hypothesis—which seems to have been endorsed by Darwin (1872/1998), Andrew (1963), and Blurton Jones and Konner (1971)—is that surprise instigates visual orienting and that the surprise display is the result or by-product of the latter (i.e., surprise $\rightarrow$ visual orienting $\rightarrow$ facial display). This latter formulation of the visual-orienting hypothesis is more accurately classified as a variant of componential theories of facial expression (e.g., Ortony & Turner, 1990; Smith & Scott, 1997), according to which the different components of facial expressions are partly controlled by separate processes, not all of which are necessarily emotional in character. According to this formulation of the visual-orienting hypothesis, surprise is but one of the conditions that instigate visual search, and the resulting facial display is not fundamentally different from similar displays that occur as the result of visual orienting due to other causes, such as when people are required to quickly look up (Reisenzein & Studtmann, 2006; see also Bennett et al., 2002; Camras, Lambrecht, & Michel, 1996).

The modifications of APT tested in our studies are not the only possible ones. However, at least two further conceivable modifications of APT can already be eliminated on the basis of our results (as well as those of previous studies): that the onset of the surprising event must be sudden rather than gradual and that the surprising event must be novel rather than familiar. All of the surprising events staged in our studies had a sudden onset, and all were novel in two salient senses of this word: First, they had not occurred before in the experiment; and second, they caused the revision of existing, and thus the acquisition of, novel beliefs (see Ruffman & Keenan, 1996).

Two other possible modifications of APT still need to be tested more thoroughly: that unexpected events, in addition to causing surprise, must also be pleasant or unpleasant rather than hedonically neutral and/or that they must be important to people's goals or welfare. The hedonic hypothesis is thrown into doubt by the findings of Experiments 6 and 7, in which surprise was coupled with high amusement. However, none of the surprising events staged in our studies were probably very important to the participants' goals or welfare.

However, note that even if future research were to support this or some other as-yet untested modification of APT, this would not change the conclusion that, in contrast to the orginal formulation of APT, surprise displays are elicited only by a (possibly small) subset of the events which cause surprise. And of course it could turn out that even the proposed modification of APT (surprise + absence of inhibition + factor X $\rightarrow$ expression) is untenable, because there is no factor X. In this case, at the latest, proponents of APT could argue that at second thought, surprise should be excluded from the domain of applicability of APT. Indeed, as mentioned in the introduction, it is possible that APT holds true for some emotions but not for others. Note, however, that this move is also not without problems. In particular, it is unsatisfactory without an explanation of why surprise does not fit APT. Given that the standard definition of basic emotions as appraisal-induced biological response syndromes (cf. the introduction) does not suggest a straightforward difference between surprise and other emotions, such an explanation may not be easy to come by.

Implications of the Dissociation Results

Regardless of how the observed dissociation between surprise and its facial expression is ultimately explained, the mere existence of this dissociation has important theoretical and practical implications. To conclude the article, we briefly mention two of them.

First, the present findings speak against any strong version of the facial feedback theory of emotional experience in the case of surprise: that facial feedback is necessary for the feeling of surprise, or that it is a major determinant of this feeling (e.g., Izard, 1977; Laird & Bresler, 1992). Because facial expressions of surprise typically did not occur in our participants, they could not have influenced their feeling of surprise.

Second, on a more practical level, our findings indicate caution in using facial expression to diagnose surprise in both research and applied settings. Perhaps the presence of a facial surprise display, or of components of that display, reliably indicates surprise in many situations (see also Reisenzein, 2000a), although certainly not in all (e.g., Camras et al., 1996; Ekman, 1979; Reisenzein & Studtmann, 2006). However, our findings suggest that, even when suppression or masking are not at work, the reverse does not hold: The absence of a facial display is no strong reason to infer a lack of surprise.

References

Andrew, R. J. (1963). The evolution of facial expression. *Science, 141,* 1034–1041.

Aramideh, M., Ongboer de Visser, B. W., Devriese, P. P., Bour, L. J., & Speelman, J. D. (1994). Electromyographic features of levator palpebrae superioris and orbicularis oculi muscles in blepharospasm. *Brain, 117,* 27–38.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. J. T. Caccioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge, United Kingdom: Cambridge University Press.

Bennett, D. S., Bendersky, M., & Lewis, M. (2002). Facial expressivity at 4 months: A context by expression analysis. *Infancy, 3,* 97–113.

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.

Camras, L. A., Lambrecht, L., & Michel, G. F. (1996). Infant "surprise" expressions as coordinative motor structures. *Journal of Nonverbal Behavior, 20,* 183–195.

Camras, L. A., Meng, Z., Ujiie, T., Dharamsi, S., Miyake, K., Oster, H., et al. (2002). Observing emotions in infants: Facial expression, body behavior, and rater judgments of responses to an expectancy-violation event. *Emotion, 2,* 179–192.

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72,* 164–176.

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.

Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior, 15,* 141–154.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91–115). New York: Guilford Press.

Darwin, C. (1998). *The expression of the emotions in man and animals* (Edited by Paul Ekman). London: Fontana Press (Original work published 1872).

Deckers, L. (1993). On the validity of a weight-judging paradigm for the study of humor. *Humor: International Journal of Humor Research, 6,* 43–56.

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: University of Nebraska Press.

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 169–248). Cambridge, United Kingdom: Cambridge University Press.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392.

Ekman, P. (1997). Expression or communication about emotion. In N. L. Segal, G. E. Weisfeld, & C. C. Weisfeld (Eds.), *Uniting psychology and biology: Integrative perspectives on human development* (Vol. 48, pp. 384–392). Washington, DC: American Psychological Association.

Ekman, P. (1999). Facial expressions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 301–321). New York: Wiley.

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

Ekman, P., Friesen, W. V., & 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.

Ekman, P., Friesen, W. V., & Hager, J. V. (2002). *Facial action coding system* (2nd ed.). Salt Lake City, UT: Research Nexus eBook.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53,* 712–717.

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

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.

Fernández-Dols, J.-M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J.-M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 255–274). Cambridge, United Kingdom: Cambridge University Press.

Fischer, A. H., Manstead, A. S. R., & Zaalberg, R. (2003). Social influences on the emotion process. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 14, pp. 171–201). Hove, United Kingdom: Psychology Press.

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

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view.* San Diego, CA: Academic Press.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23,* 567–589.

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2,* 45–67.

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.

Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology, 69,* 280–288.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40,* 129–141.

Hiatt, S. W., Campos, J. J., & Emde, R. N. (1979). Facial patterning and infant emotional expression: Happiness, surprise, and fear. *Child Development, 50,* 1020–1035.

Holodynski, M. (2004). The miniaturization of expression in the development of emotional self-regulation. *Developmental Psychology, 40,* 16–28.

Horstmann, G. (2002). Facial expressions of emotion: Does the prototype represent central tendency, frequency of instantiation, or an ideal? *Emotion, 2,* 297–305.

Horstmann, G., & Schützwohl, A. (1998). Zum Einfluß der Verknüpfungsstärke von Schemaelementen auf die Stärke der Überrraschungsreaktion [Effect of strength of association of schema elements on the surprise reaction]. *Zeitschrift für Experimentelle Psychologie, 45,* 203–217.

Izard, C. E. (1977). *Human emotions.* New York: Plenum Press.

Izard, C. E. (1991). *The psychology of emotions.* New York: Plenum Press.

Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (2001). Social context effects on facial activity in a negative emotional setting. *Emotion, 1,* 51–69.

Klix, F., van der Meer, E., & Preuß, M. (1984). Semantische Relationen: Erkennungsaufwand und psychophysiologische Reaktionstendenzen [Semantic relations: Recognition expenditure and psychophysiological reaction tendencies]. In F. Klix (Ed.), *Gedächtnis, Wissen, Wissensnutzung* (pp. 156–172). Berlin: Deutscher Verlag der Wissenschaften.

Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology, 59,* 107–128.

Kraut, R. E., & Johnston, E. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37,* 1539–1553.

Laird, J. D., & Bresler, C. (1992). The process of emotional experience: A self-perception theory. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 213–234). Thousand Oaks, CA: Sage.

Landis, C. (1924). Studies of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology, 4,* 447–509.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology, 40,* 776–785.

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.

Maher, T. F., & Furedy, J. J. (1979). A comparison of the pupillary and electrodermal components of the orienting reflex in sensitivity to initial stimulus presentation, repetition, and change. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 381–391). Hillsdale, NJ: Erlbaum.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5,* 175–190.

Meyer, W.-U., & Niepel, M. (1994). Surprise. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 353–358). Orlando, FL: Academic Press.

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion, 5,* 295–311.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion, 21,* 251–274.

Niepel, M. (2001). Independent manipulation of stimulus change and unexpectedness dissociates indices of the orienting response. *Psychophysiology, 38,* 84–91.

Orne, M. T. (1962). On the social psychology of the psychological experiment. *American Psychologist, 17,* 776–783.

Ortony, A., & Turner, W. (1990). What's basic about "basic" emotions? *Psychological Review, 97,* 315–331.

Parkinson, P. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review, 9,* 278–311.

Parrott, W. G., & Gleitman, H. (1989). Infants' expectations in play: The joy of peek-a-boo. *Cognition and Emotion, 3,* 291–311.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Reisenzein, R. (2000a). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion, 14,* 1–38.

Reisenzein, R. (2000b). The subjective experience of surprise. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 262–279). Philadelphia: Psychology Press.

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 International Society for Research on Emotions* (pp. 292–296). Toronto, Ontario, Canada: ISRE.

Reisenzein, R., & Studtmann, M. (2006). *On the expression and experience of surprise: No evidence for facial feedback, but evidence for a reverse self-perception effect.* Manuscript under review.

Reissland, N., Shepherd, J., & Cowie, L. (2002). The melody of surprise: Maternal surprise vocalizations during play with her infant. *Infant and Child Development, 11,* 271–278.

Rimé, B., Phillipot, P., & Cisamolo, D. (1990). Social schemata of peripheral changes in emotion. *Journal of Personality and Social Psychology, 59,* 38–49.

Rohrbaugh, J. W. (1984). The orienting reflex: Performance and central nervous system manifestations. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 323–373). Orlando, FL: Academic Press.

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition and Emotion, 8,* 201–229.

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.

Ruffman, T., & Keenan, T. R. (1996). The belief-based emotion of surprise: The case for a lag in understanding relative to false belief. *Developmental Psychology, 32,* 40–49.

Ruiz-Belda, M-A., Fernández-Dols, J.-M., Carrera, P., & Barchard, K. (2003). Spontaneous facial expressions of happy bowlers and soccer fans. *Cognition and Emotion, 17,* 315–326.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329–349.

Scherer, K. R., Zentner, M. R., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion, 4,* 389–402.

Schützwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1182–1199.

Schützwohl, A., & Krefting, E. (2001). Die Struktur der Intensität von Überraschung [The structure of the intensity of surprise]. *Zeitschrift für Experimentelle Psychologie, 48,* 41–56.

Siddle, D. A. T., & Jordan, J. (1993). Effects of intermodality change on electrodermal orienting and on the allocation of processing resources. *Psychophysiology, 30,* 429–435.

Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expression. In J. A. Russell & J. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Cambridge, United Kingdom: Cambridge University Press.

Sonnemans, J., & Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cognition and Emotion, 8,* 329–350.

Stekelenburg, J. J., & van Boxtel, A. (2002). Pericranial muscular, respiratory, and heart rate components of the orienting response. *Psychophysiology, 39,* 707–722.

Suls, J. M. (1971). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In J. H. Goldstein & P. E. McGhee (Eds.), *The psychology of humor* (pp. 81–100). New York: Academic Press.

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

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Volume I. The positive affects.* New York: Springer.

Tracy, J. L., Robins, R. W., & Lagattuta, K. H. (2005). Can children recognize pride? *Emotion, 5,* 251–257.

van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity over facial, jaw, oral, and neck muscles. *Psychophysiology, 38,* 22–34.

Wagner, H., & Lee, V. (1999). Facial behavior alone and in the presence of others. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 262–286). New York: Cambridge University Press.

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.

Wicklund, R. A., & Frey, D. (1980). Self-awareness theory: When the self makes a difference. In M. Wegner & R. R. Vallacher (Eds.), *The self in social psychology* (pp. 31–54). New York: Oxford University Press.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology.* New York: Holt.

Received April 4, 2005
Revision received January 30, 2006
Accepted February 6, 2006 ■