Surprise and Schema Strength

Achim Schützwohl
University of Bielefeld
Germany

Date: October 9, 1997

Running head: Surprise and Schema Strength

Abstract

Four experiments investigated the effects of stimuli discrepant with schemata of varying strength on three surprise components: The interruption of ongoing activities (indexed by response time increase), the focusing of attention on the schema-discrepant stimulus (indexed by memory performance), and the feeling of surprise (indexed by self-reports). Response times were consistently found to increase with schema strength. This effect was attributed to the increasing difficulty of schema revision. In contrast, memory for the schema-discrepant stimulus was not affected by schema strength, supporting the hypothesis that schema-discrepant stimuli are stored in memory with a distinct tag. Finally, self-reports of surprise intensity varied with schema strength only if they were made immediately after the surprising event without any intervening questions, suggesting that self-reports of surprise are highly susceptible to memory distortions.

The purpose of the experiments described in this article is to contribute to the empirical investigation of the emotion of surprise. More precisely, the aim of the present study was to examine the effects of the strength of the cognitive structure or schema from which the surprise-eliciting event deviates, on three components of the surprise syndrome, namely the subjective feeling of surprise, the interruption of ongoing activities and the focusing of attention on the surprise-eliciting event.

Previous research on surprise has been primarily focused on empirical demonstrations of the surprise reaction in response to schema-discrepant events and paid little attention to the characteristics of surprise-eliciting conditions and their effects on the different components of surprise addressed in the present study (e.g., Charlesworth, 1964; Desai, 1939; Meyer Niepel, Rudolph, & Schützwohl, 1991; Niepel, Rudolph, Schützwohl, & Meyer, 1994). For example, Meyer et al. (1991) showed that a schema-discrepant event indeed causes the surprise reaction typically consisting of the subjective feeling of surprise (as indexed by verbal reports), the interruption of ongoing activities (as indexed by action delay), and the focusing of attention on the surprise-eliciting event (as indexed by memory for this event). However, no efforts were made to systematically examine the potential effects of antecedent conditions on these surprise components.

The present experiments are directed at this research lacuna. They focus on the effects of one specific antecedent condition, namely the strength of the schema from which the surprise-eliciting event deviates, on the intensity of the subjective feeling of surprise, the duration of the interruption of ongoing activities, and the focusing of attention on the surprise-eliciting event. The following questions were addressed: Does (a) the reported intensity of the feeling of surprise, (b) the duration of action delay, and (c) memory performance for the surprise-eliciting event increase with increases in schema strength?

A Psychoevolutionary Model of Surprise

The theoretical framework for the present work is a psychoevolutionary model of surprise described in more detail in Meyer, Reisenzein, and Schützwohl (1995; in press) and Schützwohl and Horstmann (in press). This model constitutes an integration and elaboration of the work of previous authors (e.g., Berlyne, 1960; Charlesworth, 1969; Darwin, 1872; Desai, 1939; Descartes, 1649/1984; G. Mandler, 1984; Shand, 1914; Tomkins, 1962).[1] It conceives of surprise as a (generally) adaptive, evolutionary-based mechanism that manifests itself at the subjective level (the feeling of surprise; e.g., Shand, 1914; Weiner, 1986), the physiological level (e.g., the galvanic skin response; Siddle, 1991; Sokolov, 1990), and the behavioral level (in particular, a distinctive facial expression; an interruption of ongoing activities and a focusing of attention on the surprising event; Darwin, 1872; Ekman & Friesen, 1975; G. Mandler, 1984; Meyer & Niepel, 1994; Meyer et al., 1991; Niepel et al., 1994).

It is assumed that surprise is elicited by unexpected events, with unexpected events being those that deviate from an activated cognitive schema. According to schema theories (e.g., G. Mandler, 1984; Rumelhart, 1984; Rumelhart & Ortony, 1977), schemata are organized knowledge structures representing concepts such as situations, objects, events, and actions at various levels of abstractness. The central functions of schemata are (a) to enable the comprehension of current input and (b) the prediction of future events (Rumelhart & Ortony, 1977). In order to fulfill these functions effectively, schemata must provide a reasonably accurate account of the environment. This, in turn, requires that they are continuously monitored with respect to their compatibility with the data presently available. It is assumed that these monitoring processes run off in an automatic fashion, that is, they are "being controlled at levels below the level of conscious awareness" (Neumann, 1984, p. 256). As long as the data are compatible with the currently activated schema, there is no need for schema revision. This need may arise, however, if a discrepancy between current input and schema is encountered (see also Holyoak, Koh, & Nisbett, 1989; G. Mandler, 1984; Wagner, 1978). Such a discrepancy between schema and current input elicits surprise. Surprise is conceived of as an evolutionary old mechanism whose function it is to enable and motivate processes that serve to analyze the surprising event and, ultimately to remove the schema-input discrepancy.

The various components of the surprise syndrome are assumed to serve specific subfunctions engendered by this main goal (cf. Meyer, Niepel, & Schützwohl, 1994; Schützwohl & Horstmann, in press). Specifically, the feeling of surprise is assumed to signal the occurrence of an input-schema discrepancy and to motivate processes aimed at the removal of this discrepancy. The interruption of ongoing activities permits the focusing of attention on the surprising event, thus preparing the thoroughgoing analysis of the surprising event.

The analysis of the surprising event is assumed to comprise, in typical cases, the following subprocesses (for more details, see Meyer et al., 1995): (a) the verification of the perception or cognition of the surprising event; (b) an appraisal of the event's implications for the individual's well-being; (c) an appraisal of its relevance for current and future action, and (d) a search for the causes and potential predictors of the surprising event (Charlesworth, 1969; Hastie, 1984; Isaacs, 1930; Kamin, 1969; Meyer, 1988). Depending on the outcome of this analysis, the schema may be extended or restructured. As a consequence of such schema revisions, the individual is in a better position to anticipate and, possibly, to control future occurrences of the previously schema-discrepant event. Thus, the ultimate adaptive function of surprise is to ensure, by means of schema revision or knowledge-updating, an effective prediction and control of the environment and hence, successful individual-environment transactions (G. Mandler, 1984; Rescorla, 1988). In addition, the event-analysis and schema revision processes presumably do not run off in an automatic fashion but instead are controlled at the level of conscious awareness (G. Mandler, 1984; Meyer et al., 1995).

Schema strength

There are two structural aspects of a schema that can vary in strength: Variable constraints and the interconnections among variables. Schemata contain variables that represent the characteristic or defining attributes of a given concept. For example, the schema for 'face' has, among others, variables for eyes, ears, mouth, nose, cheeks, and chin. To facilitate the association of each schema variable with the correct situational aspect, the variables are specified by constraints which define the normal range of possible values of the variable. For instance, the constraints of the variable for eyes in the face schema specify the acceptable shape, size, color etc. of the eyes and thus help the assignment of appropriate values to this variable.

Strong variable constraints allow only a well-delineated range of possible values as schema-compatible. That is, they impose tight and rigid restrictions on what counts as schema-congruent input. In contrast, weak variable constraints are more flexible with respect to acceptable input (Rumelhart & Norman, 1978). For instance, the face schema of adults is most likely characterized by relatively tight variable constraints with respect to the acceptable shape, size, and color of the eyes, ears, mouth, and nose. In contrast, the face schema of young infants is probably much less restrictive in these respects.

Furthermore, schemata contain interconnections among variables which specify how they are related to one another. These relations can be temporal, spatial, correlational, or causal (e.g., Crocker, Fiske, & Taylor, 1984). In the example of the face schema, the relations among the variables are for the most part spatial, specifying the correct locations of eyes, ears, mouth, nose etc. The strength of the interconnections among variables ranges from obligatory to optional or very weak (J. Mandler, 1984). If the connection among two variables is obligatory, one schema element implies or at least strongly predicts the existence of the other element. For example, in adults, the observation of the eyes in an otherwise veiled face implies the existence of the unobserved elements of the face, such as ears, mouth, nose, etc. In contrast, if the connection among two variables is weak, one schema element does not imply or only weakly predicts the existence of the other element. For example, in early infancy, the interconnections among variables of the face schema are probably rather weak. As a result, young infants do not yet differentiate between regular faces and faces with spatial disarrangements or omissions of some of their constituent elements (Kagan, 1971).

The present experiments tested the effects of one aspect of schema strength on surprise, namely the strength of the variable constraints. It is assumed that the strength of the variable constraints is determined by (a) the frequency of the activation of the schema (e.g., G. Mandler, 1984; Thorndyke & Hayes-Roth, 1979) and (b) the variability of the values with which the schema variables have been associated in the past (e.g., Posner & Keele, 1968). If the variability of the variable values is low, then the variable constraints are expected to become gradually tighter and more rigid with an increasing number of schema activations (G. Mandler, 1984). In contrast, if the variability of the variable values is high, then the variable constraints should remain relaxed and flexible even if the schema has been frequently activated. Accordingly, in Experiments 1, 2, and 4, the strength of the variable constraints was manipulated by varying the frequency of schema activation prior to the presentation of the schema-discrepant event while keeping the variability of the values of the schema variables low. In contrast, in Experiment 3 the frequency of schema activation was held constant, but the variability of one of the schema variables was varied.

Objectives and Predictions of the Present Research

The experimental procedure used to study the effects of schema strength on surprise was adopted from Meyer et al. (1991; see also Niepel et al., 1994). Subjects performed a choice reaction-time task. In each trial, they had to press one of two response keys depending on the position of a dot (task-relevant stimulus) which appeared during the presentation of two distractor words (task-irrelevant stimuli). In the experimental groups, both distractors were displayed in normal video (NV) mode, that is, as black letters against a white background. In the critical trial, however, a salient change in the appearance of the distractors occurred: One of the two words was now displayed in reverse video (RV) mode, that is, as white letters against a black background. In the control groups, one of the two words was presented in RV mode in every single trial. Immediately after the critical trial, subjects had to rate the degree of the feeling of surprise elicited by the RV word in the critical trial and to recall the words presented in this trial. The recall of the RV word as compared to that of the NV word in the critical trial served as an index of the focusing of attention on the schema-discrepant event (see Meyer et al., 1991).

It was expected that the description of the stimuli and of the task provided in the instructions, together with the previous experience of subjects with the various modes of presentation of stimuli on computer screens, would lead to the activation or construction of an initially fairly vague schema that would have relaxed and flexible variable constraints concerning the appearance of the stimuli. With increasing number of schema activations in each uniform trial, the tightness of the variable constraints was expected to gradually increase (Experiments 1, 2, and 4). In Experiment 3, in contrast, the mode of presentation of the words in the experimental condition was either varied or remained constant in an equal number of trials. This procedure was expected to preserve relaxed and flexible variable constraints concerning the mode of presentation of the words in the former condition, whereas it was expected to result in tight and rigid variable constraints in the latter condition.

The following effects of these manipulations of schema strength on surprise were predicted:

1. The surprise ratings (the index of the intensity of the feeling of surprise) are higher in the experimental than in the control groups. This prediction was based on the assumption that the RV word in the critical trial should be a schema-discrepant event for experimental but not for control subjects and hence should elicit surprise only in the former groups and on the pertinent empirical findings of Meyer et al. (1991) and Niepel et. al. (1994).

Furthermore, within the experimental condition surprise ratings were predicted to increase with schema strength. This prediction was based on the combination of the following two assumptions: First, an event conflicting with a strong schema should be more strongly schema-discrepant than an event conflicting with a weak schema. Second, the degree of unexpectedness or schema-discrepancy is a prime determinant of the intensity of the feeling of surprise (e.g., Ortony, Clore, & Collins, 1988).

2. The response times (RTs) in the critical trial (the index of the duration of the interruption of ongoing activities) were predicted to be longer for experimental than for control subjects. This prediction was based on the assumption of the theoretical model of surprise described in the Introduction that schema-discrepant events interrupt ongoing activities and on the pertinent empirical findings of Meyer et al. (1991) and Niepel et. al. (1994).

Furthermore, it was predicted that RT in the critical trial would increase with schema strength. This latter prediction was based on the consideration that processes aimed at the integration of the schema-discrepant event into the existing schema (i.e., schema revision) may become increasingly more difficult, and as a consequence would need more time, with increasing schema strength (for a similar argument see Belmore, 1987; Bindra, 1959; Rumelhart, Smolensky, McClelland, & Hinton, 1986). More precisely, because the variable constraints on schema-congruent input in a weak schema are relaxed and flexible, they should be easily modifiable to adapt the schema to a discrepant input without the necessity to change the nature of the constraints (i.e., relaxed and flexible). In contrast, a schema characterized by rigid variable constraints should be more difficult to "bring into line" with the schema-discrepant input, because the variable constraints have to be transformed from tight and rigid constraints into more relaxed and flexible ones.

3. Concerning the memory for the two words in the critical trial (the index of the focusing of attention on the surprising event), it was predicted that experimental subjects would be better able to recall the RV word than the NV word in the critical trial. Furthermore, the RV word but not the NV word was predicted to be better recalled by the experimental subjects than by the control subjects (cf. Meyer et al., 1991). These predictions were based on the assumption that experimental subjects focus their attention selectively on the word presented in a schema-discrepant mode (RV word) while simultaneously neglecting the word presented in a schema-congruent mode (NV word; see also Metcalfe, 1995). Control subjects were expected to pay just as little attention to the (action-irrelevant and schema-congruent) RV word as to the NV word presented in the critical trial.

With respect to the effects of schema strength on memory for the RV word in the experimental groups, no firm predictions seemed possible. On the one hand, memory performance for the RV word might be expected to be independent of schema strength because the function of the focusing of attention is assumed to consist of preparing the subsequent analysis of the schema-discrepant event. Therefore, the focusing of attention is assumed to occur regardless of the strength of the disconfirmed schema. On the other hand, if the length of the interruption of ongoing activities caused by schema-discrepant events increases with schema strength, as predicted above, memory performance for the RV word might alternatively also be enhanced given a strong than a weak schema. Enhanced memory performance could be due to a longer duration of the focusing of attention to the event or to deeper encoding processes (Cermak & Craik, 1979; Metcalfe, 1995).

However, other relevant research suggests that the former prediction is more plausible, because the duration of capacity allocation or of encoding processes does not necessarily covary with memory performance (Graesser, 1981; Graesser, Gordon, & Sawyer, 1979; Hunt & Mitchell, 1982). For example, the schema-pointer plus tag model proposed by Graesser (1981; see also Davidson, 1994), assumes that memory performance for schema-incongruent items is better than for schema-congruent ones because incongruent items are stored in memory with a distinct tag that facilitates their retrieval, rather than because of the amount of processing resources allocated during acquisition. Similarly, the hypothesis proposed by Hunt and Mitchell (1982) to account for the superior recall of distinct items "implies no correlation between capacity allocation and memory" (Hunt & Mitchell, 1982, p. 85). In addition, as mentioned, the longer RT delays in the stronger schema conditions were assumed to be mainly due to increasing difficulties with schema revision processes rather than to a more extended analysis of the schema-discrepant RV word.

Experiment 1

In Experiment 1, schema strength was varied by manipulating the frequency of the activation of the schema prior to the presentation of the schema-discrepant event, while keeping the variability of the values of the schema variables low. Four experimental and four control groups were tested. In the four experimental groups, the RV word was first presented in Trial 3, 13, 23 or 33, respectively. In the corresponding control groups, one of the two words was presented in RV mode in every single trial.

Method

Participants. The participants were 68 male and 62 female students at the University of Bielefeld, Germany, with a mean age of 23.8 years (SD = 4.0). They were paid DM 2 for their participation.

Design. The experiment consisted of a 2 (experimental vs. control) x 4 (3, 13, 23, or 33 trials) between-subjects factorial design. The number of participants in the eight experimental groups varied between 15 and 18 subjects.

Apparatus. Stimuli were presented on an Atari SM 124 monochrome monitor controlled by an Atari 1040ST microcomputer.

Stimuli. Each trial started with the display of a black fixation cross against a white background. The display of the fixation cross was followed by the presentation of two distractor words, one word above the other. In the trials preceding the critical one, the distractor words in the experimental groups were presented in normal video mode (NV), that is, as black letters against a white background. In the critical trial, in contrast, one of the words was displayed in reverse video mode (RV), that is, as white letters against a black background. In the control groups, one word was presented in NV mode and the other word in RV mode in every single trial. The target stimulus consisted of a small black dot which appeared during the presentation of the word pair either above or below the words. Examples of the NV and RV trials along with the sequence of presentation of the stimuli are shown in Figure 1.

Insert Figure 1 about here

The fixation cross was presented for 1,400 ms at the center of the monitor. The display of the word pairs lasted for 3,000 ms. Word size on the monitor was

2.9 cm x 0.4 cm and distance between the two words was 0.8 cm. The words were six-letter German nouns.

The dot appeared for 100 ms either 0.4 cm above the middle of the upper word or 0.4 cm below that of the lower word. The stimulus onset asynchrony (SOA) between word and dot display was fixed at 500 ms. The next trial began 900 ms after the offset of the words or after the response, whichever occurred last. During the interval between word offset and the presentation of the next fixation cross, the computer screen remained blank.

Procedure. The participants were tested in groups of two, with the two participants who were tested together always being assigned to the same experimental condition. They were seated back to back approximately 1 m apart and wore soundproof headphones.

The experiment began with written instructions that were presented on the screen. The participants were provided with a description of the stimuli and the task. They were told to press one of the two labeled mouse keys of the computer as quickly as possible, depending on the location of the dot. If the dot appeared above the upper word, the participants were to press the left key with their right index finger. If the dot appeared below the lower word, they were required to press the right key with their right middle finger. The participants in the control groups were additionally informed that in each trial, one word would be presented in NV mode and the other one in RV mode. Any remaining questions were answered by the female experimenter at this point.

The participants then performed two practice trials where the distractor words were always presented in NV mode in the experimental groups. In the control groups, one word was presented in NV mode and one in RV mode in each of the two practice trials. Subsequently, the participants were reminded to respond as quickly as possible and then started the experimental trials by pressing a key. Depending on experimental condition, the critical trial was preceded by either 2, 12, 22, or 32 trials. Position of the RV word and of the dot in the critical trial were orthogonally combined, with each combination being presented approximately equally often within each condition.

Immediately after the critical trial, the participants were asked to complete a four-item questionnaire (in this order): (a) "Do you remember one or both of the words in the last word pair? If yes, please write it down." (b) "Did the dot appear above or below the last word pair?" (above, below, don't know). (c) "Was there anything unexpected about the displays on the screen? If yes, what was it?" (two lines were provided for the participants' free responses). (d) "How surprising was the unexpected event for you?" A 9-point rating scale was provided with the end-points "not at all" (0) and "very strongly" (8).

Results and Discussion

Unexpected event and surprise ratings. Two participants in the experimental groups failed to complete the questionnaire, answering only the recall question concerning the two words presented in the critical trial. As predicted, none of the 66 participants in the control groups referred to the RV word as unexpected. In contrast, of the 62 participants in the experimental groups, 51 mentioned the appearance of the RV word in the critical trial as an unexpected event, χ^2 (1; N = 128) = 86.9, $p < .001$. Furthermore, as shown in Table 1, the number of trials preceding the critical one had no effect on the frequency with which the RV word was mentioned as an unexpected event in the experimental groups, χ^2 (3; N = 62) = 1.1.

Insert Table 1 about here

The mean surprise ratings are also summarized in Table 1 (participants who did not make a rating were assigned a value of 0). A two-way between-subjects analysis of variance (ANOVA) of the surprise ratings with group (experimental vs. control) and number of trials (3, 13, 23, or 33) as factors, revealed only a reliable main effect for group, $F(1, 120) = 53.4$, $MSE = 4.5$, $p < .001$. As predicted, the mean surprise rating was significantly higher for experimental (3.8) than for control subjects (1.1). The other main effect and the interaction were not significant, Fs < 1.4. A separate one-way ANOVA of the surprise ratings in the experimental condition with number of trials as the factor also revealed no reliable effect, $F < 0.5$.

Thus, in accord with the predictions the presentation of the RV word in the critical trial was regarded as an unexpected event that elicited surprise in the experimental groups but not in the control groups. However, contrary to predictions, the surprise ratings did not significantly differ among the four experimental groups. Possible reasons for this unexpected finding will be considered in the Introduction to Experiment 4.

Response times. False responses and RTs exceeding the respective group mean by more than three standard deviations were treated as missing values in the RT analyses in all experiments reported here. In Experiment 1, two false responses were registered in the experimental groups and six in the control groups in the two trials preceding the critical one. Five false responses occurred in the critical trial in the experimental groups and five in the control groups. The RTs of one control subject (3-trials group) and three experimental subjects (one in each of the 3-, 13- and 33-trials group, respectively) exceeded the three standard deviations criterion. A two-way between-subjects ANOVA of the remaining subjects' average RT of the two trials preceding the critical one (Baseline), with group and number of trials as factors, yielded no reliable effects, Fs < 1.

The mean differences between the RTs in the critical trial and the Baseline are shown in Figure 2. A four-way between-subjects ANOVA with group (experimental vs. control), number of trials (3, 13, 23, or 33), position of the RV word (upper vs. lower) and position of the dot (above vs. below) revealed reliable main effects for group, $F(1, 80) = 63.1$, $MSE = 15{,}293$, $p < .001$, and number of trials, $F(3, 80) = 7.0$, $p<.01$. More importantly, the predicted Group X Number of trials interaction was also reliable, $F(3, 80) = 3.2$, $p < .05$. Planned comparisons of the RT differences between adjacent experimental groups, using t-tests, revealed significant differences for the comparisons between the 3-trials and the 13-trials groups (30 ms vs. 157 ms), $t(25) = 2.3$, $SE = 53.9$, $p < .05$, and the 13-trials and 23-trials groups (157 ms vs. 269), $t(27) = 1.8$, $SE = 61.7$, $p < .05$ (one-tailed). In contrast, the comparison of the RT differences of the 23-trials and 33-trials experimental groups was not significant (269 ms vs. 245 ms), $t(26) = 0.4$. Planned comparisons for the control groups yielded no significant effects, ts < 1.4.

Insert Figure 2 about here

In addition, the four-way ANOVA also yielded a reliable interaction between group and position of the RV word, $F(1, 80) = 5.8$, $p < .05$, indicating that RT increase in the critical trial was longer for experimental subjects when the RV word was presented in the lower than in the upper position (205 ms vs. 152 ms), whereas control subjects showed a small RT decrease when the RV word appeared in the lower position (-8 ms) and in the upper position (-13 ms). The remaining main effects and interactions were not reliable, Fs < 2.6. Finally, a comparison of the RTs in the critical trial with the Baseline within each of the four experimental and four control groups revealed significant RT increases only for the three experimental groups with 13, 23, and 33 trials, ts > 3.4, ps < .01.

The RT increases in the four experimental groups is in line with the assumption that processes of schema revision are more difficult and require more processing resources the greater the strength of the disconfirmed schema, up to a point. This point seemed to be reached in the present experiment after 23 uniform trials, given that the RT increase in the 33-trials group did not differ from that of the 23-trials group.

Recall. Table 1 shows the percentage of the participants in the experimental and control groups who recalled the RV and the NV word and the position of the dot in the critical trial. As predicted, experimental subjects were better able to recall the RV than the NV word (39% vs. 8%), $\chi^2(1, N = 64) = 17.4$, $p < .001$, whereas control subjects' recall performance for the RV and the NV word was almost identical (8% vs. 6%), $\chi^2(1, N = 66) = 0.1$. Furthermore and also in accord with the predictions, the experimental subjects differed reliably from control subjects with respect to the recall of the RV word, $\chi^2(1, N = 130) = 18.6$, $p < .001$, but not with respect to the recall of the NV word, $\chi^2(1, N = 130) = 0.2$. Furthermore, as predicted, the number of trials had no influence on the recall of the RV word by experimental and control groups, χ^2s < 2.0.

The position of the dot in the critical trial was correctly recalled by 81% of the experimental and by 47% of the control subjects, χ^2 (1; $N = 128$) = 14.2, $p < .01$. For experimental subjects, recall of the dot position in the critical trial was also independent of the number of trials, $\chi^2(3; N = 62) = 1.7$. However, control subjects with 3 and 13 trials recalled the dot's position better than control subjects with 23 and 33 trials (66% and 71% vs. 20 % and 31%), χ^2 (3; $N = 66$) = 11.2, $p < .05$.

The experimental subjects' better recall of the RV word than that of the NV word in the critical trial supports the assumption that the schema-discrepant RV word but not the schema-congruent NV word received attention (see also Metcalfe, 1995). This assumption is additionally supported by the poor recall performance of the control groups for both the RV and the NV word in the critical trial, which indicates that the action-irrelevant and schema-congruent words received only little attention even in the early trials of the experiment. Furthermore, the fact that recall of the RV word did not improve with schema strength and thus with the duration of the interruption of ongoing activities is in agreement with the schema-pointer plus tag model (Graesser, 1981) and with the "distinctiveness hypothesis" proposed by Hunt and Mitchell (1982). Both assume that memory performance does not necessarily covary with the duration of capacity allocation or of encoding processes. In addition, the fact that the recall of the RV in the critical trial was independent of the number of preceding trials is in line with the assumption that the more pronounced RT increase in the stronger schema groups is mainly due to differences in the processes of schema revision, rather than to a more extended analysis of the schema-discrepant RV word.

Experiment 2

A main result of Experiment 1 was the differential RT increases in the four experimental groups in the critical trial. It was argued that these differential RT increases reflect differences in the difficulty of the processes aimed at the integration of the schema-discrepant event into the existing schema. To corroborate this interpretation, a second experiment was conducted to provide empirical evidence that the occurrence of a schema-discrepant event indeed instigates schema revision processes. The main prediction tested in Experiment 2 was that subsequent presentations of a RV trial would be considerably less schema-discrepant or even not schema-discrepant at all, regardless of the strength of the disconfirmed schema. Consequently, subsequent RV trials should lead to a smaller or no RT increase. This prediction was based on G. Mandler's (1984) assumption that whereas "the normal [unconscious] course of learning is slow and cumulative" with the intervention of conscious awareness, schema revision can be "established quickly and in a saltatory fashion" (G. Mandler, 1984, p. 106).

It is interesting to compare this schema theoretical approach with the predictions of neural network models of learning which have been offered as alternatives to schema theory (e.g., McClelland & Rumelhart, 1988; Rumelhart, Smolensky, McClelland, & Hinton, 1986). The reason is that although these neural network models may be able to explain the RT findings in the single RV trial in the experimental groups of Experiment 1, it appears that the prediction derived from simple neural network models concerning the RTs in subsequent RV trials is at variance with the schema theoretical prediction described above.

To see this, let me first consider a possible neural network explanation of the RT findings in the RV trial in the experimental groups of Experiment 1. According to these models, learning consists of the modification of the weights or the strength associated with the connections between the units of a neural net. Because the learning rate is a fixed parameter set at a low level, learning should always be a slow and continuous process. Hence, in Experiment 1 the strength of the connections among the units of the net should increase slowly and continuously with every NV-trial (up to the asymptote of learning) preceding the critical RV trial. To account for the RT delays in the critical trial, one would have to additionally assume, for example, that a neural net with strong connections among its units needs longer to settle in a stable state when disturbed (critical RV trial) than a net with weaker connections and that a response is emitted only after the net has reached that stable state.

In contrast, the predictions derived from the schema-theoretical and the neural network approach concerning RTs in subsequent RV trials diverge. Because simple neural networks do not change the learning rate in a saltatory fashion and hence always learn slowly, they would seem to make the following predictions: (a) The time to settle in a stable state, and hence the RTs in the RV trials following the critical trial should initially be above Baseline and then decrease continuously to the Baseline. (b) RTs in these trials should be longer, and the number of trials necessary to return to the Baseline should be higher, the more uniform trials preceded the critical one and hence the stronger the connections between the network units before the first RV trial.

In Experiment 2, the same four experimental groups as in Experiment 1 were tested, but no control groups were included. However, to examine the effects of the first RV display on subsequent RV presentations, the first RV trial was immediately followed by nine additional RV trials. After the tenth RV trial, a final NV trial was presented to test whether the re-presentation of the NV word after ten RV trials would now constitute a schema-discrepant event for the participants.

Method

Participants. The participants were 43 female and 37 male students at the University of Bielefeld, Germany, with a mean age of 24.2 years (SD = 4.4). There were 20 participants in each of the four experimental groups. They were paid DM 2 for their participation.

Apparatus and stimuli. Apparatus and stimuli were the same as in Experiment 1.

Procedure. The procedure in Experiment 2 was the same as in Experiment 1 with the exception of the following changes: (a) Only four experimental but no control groups were tested. (b) In each of the four experimental groups, the critical trial (Trial 3, 13, 23, or 33, respectively) was followed by nine additional RV trials. The positions of the RV word and of the dot in the ten RV trials were balanced in each group. (c) In the critical trial, the RV word appeared always in the lower position and the dot appeared always above the upper NV word. (d) The final trial in each group consisted of a NV trial.

The same questionnaire as in Experiment 1 was used. However, in contrast to Experiment 1, it was not presented immediately after the critical trial, but only at the end of the experiment, that is, after ten more trials. In addition, the questions concerning the recall of the words and the position of the dot did not refer to the critical trial but to the last trial (i.e., the final NV trial). The reason for administering the questionnaire only at the end of the experiment was that the RTs in the trials following the critical one were of main interest. Therefore, it was deemed advisable not to risk contaminating these data by processes elicited by answering the questionnaire.

Results and Discussion

Unexpected event and surprise ratings. Only 35% of the participants in the 3-trials group but at least 85% of the participants in each of the three remaining groups considered the RV-presentation an unexpected event, χ^2 (3; N = 80) = 20.9, $p < .001$. In addition, one participant mentioned the re-presentation of the NV word in the last trial after ten RV trials as unexpected. A one-way ANOVA of the surprise ratings resulted in a significant effect of the group factor, $F(3, 76) = 8.2$, MSE = 4.0, $p < .001$. A subsequent multiple range test (Duncan, $p < .05$) revealed that the mean surprise rating in the 3-trials group (0.90) was reliably lower than the mean surprise ratings in the 13, 23, and 33-trials groups (3.45, 3.35, and 3.55, respectively). Thus, in contrast to Experiment 1, the number of trials preceding the critical trial influenced the frequency of mentioning the RV event as unexpected and the surprise ratings.

However, these findings may have been due to the fact that the questionnaire was not administered immediately following the critical trial but only after nine more RV trials and one final NV trial. As a consequence of this procedure, in the 3-trials group there were altogether ten RV trials but only three NV trials before the questionnaire was answered. In contrast, in the other groups there were always more NV trials than RV trials. Due to this difference in the ratio between RV and NV trials the subjects in the 3-trials group may not retrospectively have considered the RV presentation as an unexpected but as a normal, because more frequent, event.

Response times. Four false responses occurred in the critical trial, and seven in the two preceding (baseline) trials. The corresponding RTs were treated as missing values. In the critical trial, no RT exceeded the respective group mean by more than three standard deviations.

A one-way ANOVA of the average RTs in the two baseline trials unexpectedly yielded a reliable group effect, $F(3, 67) = 5.7$, MSE = 14,890, $p < .01$. This effect was due to the fact that the Baseline was more than 100 ms longer in the 13-trials group (545 ms) than in each of the remaining groups (419 ms, 438 ms, and 391 ms, respectively).

The mean differences between the RT in the critical trial and the Baseline are shown in the upper panel of Figure 3. A one-way ANOVA of the differences revealed a reliable group effect, $F(3, 67) = 3.9$, MSE = 53,660, $p < .05$. Planned comparisons of the RT differences between adjacent groups, using t-tests, revealed only a significant difference between the 3-trials and the 13-trials group (94 ms vs. 244 ms), $t(35) = 2.6$, SE = 57.8, $p < .05$, whereas the comparisons between the 13-trials and the 23-trials group and between the 23-trials and 33-trials group were not significant, ts < 1.3. Following a suggestion by Fazio (1990) concerning the treatment of unequal baselines, a percentage RT increase was computed (the ratio of RT differences in the critical trial to baseline RT). These data are shown in the lower panel of Figure 3.

Insert Figure 3 about here

A one-way ANOVA of the percentage RT increase was significant, $F(3, 67) = 4.6$, MSE = 3,310, $p < .01$. Planned comparisons of the percentage RT increases between adjacent groups fully replicated the results obtained in

Experiment 1: The 3-trials group differed significantly from the 13-trials group (24% vs 50%), $t(35) = 2.2$, SE = 12.2, $p < .05$, and the 13-trials from the 23-trials group (50% vs. 83%), $t(36) = 1.9$, SE = 17.4, $p < .05$ (one-tailed). As in Experiment 1, the 23-trials group and the 33-trials group did not differ from each other (83% vs. 86%), $t(32) = 0.1$.

In a subsequent analysis step, the Baseline RT was compared with the RT in the critical trial within each of the four groups. These comparisons revealed, in each group, significant RT increases from the Baseline to the critical trial, ts > 4.0, ps < .01.

To examine the effects of the first RV presentation on the RTs in subsequent RV trials, the RTs in the two trials following the critical trial were separately compared with the Baseline and with the RT in the critical trial within each of the four groups. RTs in the two trials after the critical trial did not significantly differ from Baseline in any of the groups, ts < 1.5. In contrast, RTs in the first and the second trial after the critical one were reliably faster than RTs in the critical trial (first trial: ts > 3.4, ps < .01; second trial: ts > 2.6, ps < .05). Furthermore, one-way ANOVAs of the percentage RT changes in the two trials after the critical one relative to the Baseline revealed no reliable differences between the four groups, Fs < 0.6.

Hence, the RTs in the RV trials following the critical trial confirmed the prediction based on schema theory and contradict the prediction derived from simple neural network models of learning. The finding that RTs in the two trials immediately following the critical one (a) were reliably faster than RTs in the critical trial, (b) no longer differed from the Baseline in each of the four groups and (c) did not differ between the groups is in line with the schema-theoretical prediction that, regardless of schema strength, the RV event would become completely integrated into the schema, as a consequence of which subsequent RV presentations would be entirely schema-congruent. The issue whether more sophisticated neural network models could provide for an alternative to the schema-theoretical approach will be discussed in the General Discussion.

Finally, the re-presentation of two NV words in the final trial did not influence RT in this trial as compared with the average RT of the two preceding trials in any of the groups, $\underline{t}$s < 0.7, confirming that subjects did not consider it an unexpected event.

Recall. In agreement with participants' statement that the re-presentation of two NV words in the final trial was not unexpected, none of them was able to recall a word presented in the final (NV) trial. The position of the dot in the final (NV) trial was recalled by 25% of the participants each in the 3-trials and 33-trials groups, and by 35% of the subjects each in the 13-trials and 23-trials groups, χ^2 (3; $\underline{N}$ = 80) = 0.9.

Experiment 3

In Experiments 1 and 2, schema strength was varied by manipulating the number of uniform trials preceding the schema-discrepant event. To generalize the findings of these experiments, Experiment 3 varied schema strength by manipulating the variability of the values with which schema variables are associated, while keeping trial number constant. Specifically, the variable constraints concerning the mode of word presentation were manipulated by either varying or keeping constant the font-types of the NV words in an equal number of trials preceding the critical one.

In one experimental group (low schema strength), the font-type of one of the two NV words was varied in each of the noncritical trials, whereas in the second experimental group (high schema strength), the font-type of both words was kept constant across the same number of trials. In the latter experimental group, two NV words were presented in the standard font-type already used in the preceding experiments in each one of 35 trials preceding the critical one. In contrast, in the low schema strength group, one of the two NV words was shown in the standard font-type and the other word in one of seven different font-types. Altogether, each of the seven font-types appeared five times in trials 1 - 35. In Trial 36, one word (that shown in standard font-type) was presented in RV mode. In the two corresponding control groups, the font-type was also either varied or kept constant. Additionally, one word (that shown in standard font-type) was presented in RV mode in each trial. Immediately after the critical trial, the subjects were asked to complete the same questionnaire already used in the preceding experiments.

Based on the results of the previous experiments it was expected that the RV word in the critical trial would be a schema-discrepant event that elicits surprise in both experimental groups but not in the control groups, and would therefore result in an increase in RT and in enhanced recall of the RV word but not of the NV word in the former groups. More importantly, RT increase was expected to be more pronounced in the experimental group who had always seen the words presented in the same font-type (high schema strength) than in the experimental group with varied font-type (low schema strength). This prediction was based on the assumption that schema revision would be more difficult and hence more time-consuming in the high schema strength group (i.e., that with tight and rigid variable constraints) than in the low schema strength group (that with relaxed and flexible variable constraints). Based on the recall performance obtained in Experiment 1, no differences concerning recall of the RV word in the critical trial were predicted between the two experimental groups.

Method

Participants. Participants were 50 female and 30 male students at the University of Bielefeld, Germany, with a mean age of 24.1 years ($SD = 4.0$). They were paid DM 2 for their participation.

Apparatus. The apparatus was identical to the one used in Experiment 1.

Stimuli. Word size was 2.9 cm x 0.4 cm for the standard font-type, whereas the words presented in the seven different font-types were between 2.7 and 3.2 cm wide and letter height varied between 0.4 and 0.6 cm. The sequence of events was the same as in the preceding experiments.

Design. The experiment consisted of a 2 (experimental vs. control) x 2 (constant font-type vs. varied font-type) between-subjects factorial design. There were 24 participants each in the two experimental groups, and 16 participants each in the two control groups.

Procedure. The procedure was the same as in Experiment 1 with the exception of two changes. First, participants were tested individually. Second, due to an error in the computer program controlling the experiment, the question concerning the position of the dot in the last trial was presented on the computer monitor immediately after the critical trial. Hence, only after participants had read this question were they handed the questionnaire.

Results and Discussion

Unexpected event and surprise ratings. None of the participants in the control groups but 77% of the participants in the experimental groups mentioned the RV word in the critical trial as an unexpected event, χ^2 (1; $N = 80$) = 42.8, $p < .001$. As predicted, experimental groups with constant and varied font-type did not reliably differ from each other in this respect (83% vs. 71%), χ^2 (1; $N = 48$) = 0.5.

A two-way between-subjects ANOVA of the surprise ratings, with groups (experimental vs. control) and font-type (constant vs. varied) as factors, yielded a reliable main effect for groups only, $F(1, 76) = 19.1$, MSE = 5.8, $p < .001$, indicating higher surprise ratings for experimental than for control subjects (3.8 vs. 1.3); remaining Fs < 2.2. A direct comparison of the surprise ratings of the two experimental groups revealed no significant difference (4.2 vs 3.3), $t(46) = 1.3$, SE = 0.7. Hence, as in the previous experiments surprise ratings did not differ significantly between experimental groups with low and high schema strength.

Response times. In the two baseline trials, only one false response (experimental condition) was registered. In the critical trial, four false responses occurred, one in the control and three in the experimental condition. RTs in the critical trial of one control subject (font-type constant) and two experimental subjects (one with constant and one with varied font-type) exceeded their respective group means by more than three standard deviations.

A two-way between-subjects ANOVA of the mean RTs of the two trials immediately preceding the critical one (Baseline), with group (experimental vs. control) and font-type (constant vs. varied) as factors, revealed no reliable effect, Fs < 1.2.

Insert Figure 4 about here

The mean differences between the RTs in the critical trial and the Baseline are shown in Figure 4. A four-way between-subjects ANOVA with group, font-type, position of the RV word (upper vs. lower) and position of the dot (above vs. below) yielded reliable main effects for group, $F(1, 57) = 29.7$, $MSE = 21{,}539$, $p < .001$, font-type, $F(1, 57) = 8.6$, $p < .01$, and position of the dot, $F(1, 57) = 4.6$, $p < .05$. The Group X Font-type interaction as well as the Font-type X Position of the dot interaction were also reliable, $F(1, 57) = 6.0$, $p < .05$, and $F(1, 57) = 5.0$, $p < .05$, respectively. The interaction between group and position of the dot was marginally significant, $F(1, 57) = 3.7$, $p < .10$. The remaining main effect and interactions were not reliable, Fs < 1.5.

A follow-up analysis of the Group X Font-type interaction revealed that RT differences were not reliably different between the control groups with constant versus varied font-type (22 ms vs. 7 ms), $t(28) = 0.8$. However, as predicted and replicating the results of the previous experiments with a different manipulation of schema strength, RT increase was reliably longer for experimental subjects with constant than for those with varied font-type (270 ms vs. 111 ms), $\underline{t}(41) = 2.7$, SE = 59.7, $\underline{p} < .05$. Note that the schema-discrepant event concerns the word presented in the constant font-type used in trials 1 - 35 in both experimental groups. That is, although it was not the varied font-type word that was presented in RV mode in the low schema strength group, RT increase was significantly less pronounced due to the variation of the font-type of the other word in the preceding trials.

This seems in good correspondence with the schema revision hypothesis proposed to account for the differential RT increases in the critical trial: The variation of the font-type of one of the two words in each trial preserves the relaxed and flexible variable constraints for the mode of word presentation. As a consequence, the variable constraints can be easily modified to adapt the schema to a discrepant input without the necessity to change the nature of the constraints (i.e., relaxed and flexible). The Font-type X Position of the dot interaction was due to a longer RT increase with constant than with varied font-type when the dot appeared below the words, (232 ms vs. 69 ms), $\underline{t}(33) = 2.2$, SE = 74.8, $\underline{p} < .05$, whereas no differences in RT increase were obtained between constant and varied font-type when the dot appeared above the words (97 ms vs 71 ms), $\underline{t}(36) = 0.7$.

Furthermore, RT increases were reliably longer in each experimental group than in its corresponding control group, $\underline{t}(35) = 3.9$, SE = 63.0, $\underline{p} < .001$, for groups with constant font-type (270 ms vs. 22 ms), and $\underline{t}(34) = 2.8$, SE = 36.5, $\underline{p} < .01$, for groups with varied font-type (111 ms vs. 7 ms). Finally, the RT increases from the Baseline to the critical trial were significant in both experimental groups, $\underline{t}(20) = 5.0$, SE = 53.7, $\underline{p}<.001$, and $\underline{t}(21) = 4.0$, SE = 27.6, $\underline{p}$

$< .01$, for constant and varied font-type, respectively, but for neither of the control groups, ts < 1.6.

Recall. Replicating the findings of Experiment 1, experimental subjects were better able to recall the RV word than the NV word in the critical trial (29% vs. 2%), χ^2 (1; $N = 48$) = 13.4, $p < .001$, whereas control subjects' recall performance for the RV and the NV word was identical (13% each), χ^2(1, $N = 32$) = 0. Furthermore, the difference of the RV word recall (29% vs. 13%) and the NV word recall (2% vs. 13%) between the experimental and the control subjects were marginally reliable, χ^2 (1; $N = 80$) = 3.1, $p < .10$, and χ^2 (1; $N = 80$) = 3.6, $p < .06$, respectively. In addition, there were no reliable differences in RV word recall (25% vs. 33%) and NV word recall (0% vs. 4%) between the experimental groups with constant and varied font-type, χ^2s < 1.0. In the control groups, both the RV word and the NV word were each recalled by 19% of the subjects with constant font-type and by 6% of the subjects with varied font-type, χ^2s < 1.2. Hence, the pattern of the recall performance for the RV and NV word in the critical trial for the most part confirms the assumption that the experimental subjects focus their attention selectively on the schema-discrepant RV word while simultaneously neglecting the schema-congruent NV word in the critical trial. Furthermore, the control subjects again seem to pay only little attention to both words.

Finally, only four experimental and four control subjects failed to correctly recall the dot's position in the critical trial, χ^2 (1; $N = 80$) = 0.1. This finding was most likely due to the fact that in Experiment 3, in contrast to Experiment 1, the question concerning the position of the action-relevant dot appeared on the monitor immediately after the response in the last trial, thus favoring recall of the dot position.

Experiment 4

Experiment 4 served two purposes. The first was to further investigate the relationship between schema strength and the feeling of surprise. Contrary to predictions, the previous experiments consistently found that the surprise ratings did not reliably vary with the manipulation of schema strength. However, this negative finding may have been due to insensitive measurement methods. For example, it could be argued that considering the short duration of surprise, the ratings were collected too late or were influenced by the answering of intervening questions. In support of this, the surprise ratings in the 3-trials group of Experiment 2, which were collected only after ten additional trials, were considerably lower than in the 3-trials group in Experiment 1. Even in Experiments 1 and 3, where surprise was rated immediately after the critical trial, participants first answered the questions concerning the words and the position of the dot in the last trial and the occurrence of an unexpected event.

To remedy these potential problems, in Experiment 4 the surprise ratings were collected immediately after the occurrence of the schema-discrepant event without any intervening questions being asked. In addition, Experiment 4 was conducted with a larger sample of subjects per cell. Finally, following Frijda, Ortony, Sonnemans, and Clore (1992), the participants answered questions concerning two other aspects of the subjective experience of surprise apart from overall intensity, namely the duration and the peak amplitude of surprise.

The second purpose of Experiment 4 was to obtain some information about the processes underlying the differential RT increases in the low and high schema strength groups observed in the previous experiments. It was previously argued that these RT differences reflect differences in the processes of schema revision. However, at present, alternative explanations cannot be excluded. For example, it could be argued that differences in one or more of the event-analysis processes proposed in the model described in the Introduction (verification, well-being, action-relevance checks, and causal analysis) were, alone or in combination, responsible for the RT results. To gain information about the processes underlying the differential RT increases, part of the participants in Experiment 4 were asked to estimate retrospectively how long they had been engaged in the various processes of the analysis of the schema-discrepant event and in schema revision.

This verbal report method was deemed worthwhile for several reasons: (a) The event-analysis and schema revision processes elicited by the schema-discrepant event are assumed to be conscious (G. Mandler, 1984), and conscious processes should be accurately reflected in verbal reports (Ericsson & Simon, 1993; Wilson, 1994). (b) The accuracy of verbal reports about conscious cognitive processes is generally assumed to increase the shorter the duration of the task and the interval between task completion and the gathering of the verbal reports (Ericsson & Simon, 1993; Wilson, 1994). As a consequence, verbal reports about the event-analysis and schema revision processes should be especially accurate because the unexpected event was of short duration and the verbal reports were gathered immediately thereafter. (c) Verbal reports can provide information that is difficult to obtain by other means (Crutcher, 1994). (d) Finally, the verbal report measure used in the present experiment is in accord with the use of verbal reports in the field of metacognition, where verbal reports have a prominent status. As stated by Nelson (1996), "the metacognitive approach is to formulate verbal reports as meta-level statements about what is occurring at the object-level, to operationalize what is occurring at the object-level through some kind of observable criterion response, and then to assess empirically the degree of relationship between the verbal report and the criterion response" (p. 106). In the present experiment the criterion response that can be compared to the verbal reports concerning the duration of the various cognitive processes is the RT increase in the critical trial in the two experimental groups.

Method

Participants. A total of 180 students at the University of Bielefeld participated in Experiment 4. Eight participants had to be excluded from the data analyses, three participants because of computer malfunction, four because they failed to notice the inverted word and hence could not answer the questionnaire, and one participant because she did not comply with instructions. This left 84 male and 88 female participants with a mean age of 23.3 years (SD = 3.5), who were not paid for their participation.

Design. There were two groups in Experiment 4, a 3-trials group and a 33-trials group. Within the 3-trials group, 46 participants answered the feeling of surprise questionnaire, and 39 participants the analysis of the surprising event and schema revision questionnaire immediately after the critical trial. Within the 33-trials group, 47 participants answered the feeling of surprise questionnaire, and 40 participants the analysis of the surprising event and schema revision questionnaire.

Apparatus and stimuli. The apparatus and stimuli were the same as in Experiment 1.

Procedure. The procedure was also the same as in Experiment 1 with the exception of the questionnaire that participants answered immediately after the critical trial. Each question was presented on a separate page. The participants were asked to complete the questionnaire by answering one question after the other without turning back pages to consult previous answers. Two different questionnaires were used. The questions contained in the two different questionnaires are reprinted in Table 2.

Insert Table 2 about here

The participants who answered the questionnaire concerned with the feeling of surprise were asked to indicate the overall intensity of surprise and its peak amplitude on nine-point rating scales with the end points "not at all" (0) and "very strongly" (8). For the question referring to the duration of the feeling of surprise, a scale ranging from 0 to 1,200 ms in 100 ms steps was provided.[2] If the participants thought that their feeling of surprise lasted still longer, they were asked to note down its estimated duration next to the scale.

The participants who answered the questionnaire referring to the analysis of the surprising event and schema revision were asked to indicate on 11-point rating scales how long they had been engaged with each of the described processes. The scale ranged from "not at all" (0) to "very long" (10), with intermediate scale points labeled "very short" (1), "short" (3), and "long" (7).

Results and Discussion

Response times. In the two baseline trials 13 false responses occurred. In the critical trial, 18 false responses were registered, 12 in the 3-trials group and six in the 33-trials group. In addition, the RTs in the critical trial of two participants in each of the two groups exceeded the three standard deviations criterion. The baseline RTs of the remaining participants were significantly longer in the 3-trials group than in the 33-trials group (516 ms vs. 456 ms), $t(131) = 2.2$, SE = 27.2, $p < .05$. A three-way between-subjects ANOVA of the differences between the RT in the critical trial and baseline RT with number of trials (3 vs. 33), position of the RV word (upper vs. lower) and position of the dot (above vs. below) as factors revealed only a reliable main effect for number of trials, $F(1, 125) = 27.7$, MSE = 51,165, $p < .001$.[3] Replicating the basic finding of the previous experiments, RT increase in the critical trial was larger for the 33-trials group than for the 3-trials group (284 ms vs. 80 ms). The remaining effects were not reliable, Fs < 2. Finally, an analysis of RT increase from Baseline to the critical trial revealed significant increases in both experimental groups, ts > 4.0, ps < .001.

Feeling of surprise. One of the 93 subjects who received the surprise questionnaire answered only the first question. The means of the obtained ratings in the 3-trials and 33-trials groups are summarized in Table 3. T-tests revealed a significant difference between the two groups with respect to overall felt surprise, $t(91) = 4.3$, SE = 0.4, $p < .001$, reflecting that overall surprise intensity was higher in the 33-trials group than in the 3-trials group (5.4 vs. 3.6; cf. Table 3). With respect to peak amplitude and duration of surprise, the differences between groups went in the same direction, but failed to reach significance, ts < 1.7. Thus, contrary to the results of the preceding experiments, overall felt surprise was found to be reliably stronger in the 33-trials group than in the 3-trials group.

Insert Table 3 about here

A comparison of this finding with those of Experiment 1 (which allows for the closest comparison with Experiment 4) suggests that it was mostly due to the fact that the surprise ratings obtained for the 33-trials group in Experiment 4 were higher (5.4) than those obtained for the 33-trials group in Experiment 1 (4.4). In contrast, the mean surprise ratings in the 3-trials groups were almost identical in the two experiments (3.6 vs. 3.4). This finding is in accord with the hypothesis that existing group differences in Experiment 1 were leveled out because the ratings were collected only after several intervening questions. Presumably, after this delay, the participants no longer had a precise memory about the actual surprise intensity, which made them uncertain which rating scale category was most appropriate. This uncertainty, in turn, may have caused them to rely on the intermediate scale categories.

In contrast to overall surprise intensity, there were no significant differences between the 3-trials and the 33-trials groups with respect to the peak amplitude and the duration of the feeling of surprise. One possible explanation for this pattern of results would be that the peak amplitude of the surprise feeling lasts longer if the schema-discrepant event deviates from a strong schema. If so, overall surprise intensity could be higher in the strong schema group than in the weak schema group even though the peak amplitude and the duration of the surprise feeling do not differ between groups.

Another explanation of these findings might be that the rating of overall surprise intensity, but not of the duration and the peak amplitude of surprise, were influenced by factors in addition to the actual intensity of surprise. For example, in line with self-perception theory (Bem, 1972), the person's awareness of the delay in the execution of the required action caused by the schema-discrepant event may have influenced the overall surprise ratings, but not the more specific question concerning the peak amplitude and the duration of the surprise feeling (see also Ritter, 1997; Schützwohl, 1997a). Parenthetically, it should be noted that if correct, this self-perception explanation of the differential effect of the number of trials on the overall surprise intensity ratings suggests that felt surprise was not in fact more intense in the 33-trials group than in the 3-trials group.

Analysis of the surprising event and schema revision. Seventy-nine participants answered the questionnaire asking for the duration of the various event-analysis processes and with schema revision. Two of them failed to complete the questionnaire. The mean ratings in the two groups are also shown in Table 3. In line with the schema revision hypothesis, the two groups differed reliably with respect to the question asking for how long they had been engaged with processes of schema revision, $t(76) = 2.4$, $SE = 0.6$, $p < .05$. Participants in the 33-trials group reported to have been engaged longer with processes of schema revision than participants in the 3-trials group (4.2 vs. 2.8). The remaining comparisons were not reliable, ts < 1.7.

Thus, the results concerning the duration of the processes of the analysis of the schema-discrepant event and of schema revision were clear-cut: As predicted, only the duration of schema revision processes was rated higher in the 33-trials group than in the 3-trials group, hence reflecting the differences in RT increase in the two groups. This lends empirical support to the assumption that schema-discrepant events are in fact more difficult to integrate into a strong than into a weak schema (e.g., Belmore, 1987; Bindra, 1959; Rumelhart et al., 1986). In addition, it is noteworthy that the subjects in both schema strength groups indicated that the discrepancy verification took only very little time. This is not surprising, because the RV display was a very salient event, thus facilitating the verification task. Similarly, the appraisal of the implications of the schema-discrepant event for well-being was judged to be of very brief duration only. This was probably due to the fact that it is easy to identify the occurrence of the RV event as irrelevant for well-being and as therefore not requiring further processing.

Somewhat more time was reported for the action-relevance check and the causal analysis of the schema-discrepant event (Table 3). However, there were again no differences between the 3-trials and the 33-trials group. Apparently, the participants in both groups arrived equally quickly at the same conclusions, namely that the change in the appearance of the words is part of the experimental procedure (Schützwohl & Reisenzein, in press) and is irrelevant for the key-pressing task (Meyer, Reisenzein, & Schützwohl, in press).

General Discussion

The purpose of the present experiments was to study the effect of schema strength on three components of surprise, namely the feeling of surprise, the interruption of ongoing activities, and the focusing of attention.

Feeling of Surprise

The feeling of surprise was reported to have been of brief duration in the present experiments, and unless measured immediately after the critical trial without any intervening questions, it did not reliably increase with schema strength. Only when the subjects rated their feeling of surprise straight after the surprising event had occurred, was overall surprise intensity found to be more pronounced when the schema-discrepant event deviated from a strong than from a weak schema. This difference was, however, restricted to the overall surprise intensity rating, whereas no significant differences were obtained for the peak amplitude and the duration of surprise. Because factors such as the person's awareness of action delay may have contributed to the rating of overall felt surprise, the issue whether the feeling of surprise increases with schema strength (as defined in the present experiments) remains as yet not fully settled.

Interruption of Ongoing Activities

It was consistently found that the duration of the interruption of ongoing activities increased with the strength of the disconfirmed schema. This finding was taken as indicating that the schema-discrepant event is increasingly more difficult to integrate into the disconfirmed schema the greater its strength. Some empirical support for this explanation was obtained in Experiment 4, in which subjects estimated the process of schema revision to be more time consuming if the disconfirmed schema was strong rather than weak, whereas the estimated duration of the other processes concerned with the analysis of the surprising event was found not to differ significantly between the strong and weak schema group.

This is of course not intended to imply that the processes of the analysis of the schema-discrepant event always take a constant amount of time. For example, in a recent experiment (Schützwohl, 1997b) it was found that an affectively negative surprising event caused a longer interruption of ongoing activities than an affectively positive one. In this case, the subjects reported no differences with respect to the duration of the schema revision process, but the causal analysis of the surprising event was reported to have taken longer when the schema-discrepant event was affectively negative.

The present RT findings seem to be relevant to a central assumption of several theories of associative learning (Öhman, 1979; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1978; cf. Dickinson, 1980). This assumption holds that one or both elements of an association (E1 and E2) are processed increasingly longer, the more unexpected or surprising they are, with unexpectedness being a function of the associative strength between the two elements. Specifically, the greater the associative strength between E1 and E2, the greater the unexpectedness of an omission or a change in the appearance of E2 after E1 has been presented.

In line with this assumption, Siddle and Packer (Packer & Siddle, 1989; Siddle & Packer, 1987) found that RTs to a secondary probe were slower during unexpected trials, indicating that unexpected events indeed demand processing resources. However, these authors did not explicitly address the theoretically important assumption of a linear (or at least monotonic) relationship between the unexpectedness of an event and the amount of processing devoted to it. In contrast, on the condition that schema strength and associative strength can be taken to be sufficiently similar concepts, the experiments presented in this article provide some empirical support for the assumption of a linear relationship between unexpectedness and amount of processing.

Focusing of attention

Experimental subjects were better able to recall the RV word than the NV word in the critical trial. In addition, RV word recall but not NV word recall differed between experimental and control groups. Finally, recall of the RV word in the experimental groups did not vary with schema strength. This pattern of results is in line with an encoding-and-retrieval explanation of the effects of surprising events on memory. Better recall of the RV word but not of the NV word in the critical trial in experimental groups strongly suggests that the surprising event receives increased attention during encoding. However, because the duration of attention allocation to the RV event did not enhance the recall of the RV word, it seems necessary to assume that the main effect of attending to the RV event was to facilitate its retrieval. This facilitation effect may have been due to the surprising event being stored separately in memory or being marked with a tag (Graesser, 1981).

This explanation bears some similarities with the processes made responsible for the superior recall of isolated items (the so called von Restorff effect; von Restorff, 1933) or distinct items (e.g., Bruce & Gaines, 1976; Fabiani & Donchin, 1995; Schmidt, 1991), but there are also a number of important differences between surprising and isolated or distinct events. First, the isolation effect is obtained under intentional learning instructions; it usually does not occur under conditions of incidental learning. In a review of the research concerned with the isolation effect, Wallace (1965) stated that "the only case of the isolation effect occurring in the absence of intent to learn was when a single item was isolated in a manner which was relevant to the learner’s task, that is, isolation was directly related to the response required of the subject" (Wallace, 1965, p. 416). In contrast, the present experiments constituted an incidental learning situation where the surprising event was irrelevant for the response required of the subjects. Nonetheless, a superior recall of the surprising word was obtained.

Second, although Green (1956) explicitly attempted to account for the isolation effect by referring to the “surprise value“ of the isolated item, surprise is clearly not a necessary condition for the isolation effect to occur. For example, both Green himself (1958) and, more recently, Fabiani and Donchin (1995) instructed their subjects in advance as to the nature of the items (including the isolated items) that would be presented, thus reducing or even eliminating surprise. Nonetheless, an isolation effect was found in both studies.

Third, for the isolation effect to occur it is not necessary that subjects know about the status of the isolated item at the time of its presentation. As a consequence, the isolation effect is also obtained in the absence of increased attention during encoding. This was in fact already shown by von Restorff (1933) in her original experiments. For example, von Restorff presented her participants lists consisting of nine heterogeneous items, followed by nine homogeneous items and an isolated item that was presented after the first homogeneous item. That is, at the time of presentation of the isolated item, the subjects had seen nine heterogeneous items and one item from the homogeneous list (which at the time of its presentation was just another heterogeneous item). Because the status of the isolated item became apparent only after the presentation of eight additional homogeneous items, the subjects could not preferentially encode the isolated item at the time of its presentation. This is in clear contrast to the experiments reported in this article, where the status of the surprising event became known at the time of its presentation, as a consequence of which it received increased attention during encoding.

Interruption of Ongoing Activities and Focusing of Attention

Taken together, the RT increases in the critical trial and the enhanced memory performance for the RV word observed in the experiments reported here indicate that schema-discrepant events capture attention. The assumption that schema-discrepant stimuli capture attention may be considered to be a variant of the contingent involuntary orienting hypothesis proposed by Folk, Remington, and Johnston (1992, 1993).

According to this hypothesis, involuntary shifts of attention depend on the relation between specific stimulus properties and characteristics of the individual such as active goals and prior experience. Folk et al. (1993) illustrate the proposed mechanism of attentional capture by using the analogy of a thermostat "in which the set-point is programmed (off-line) on the basis of top-down goals, but the on-line response of the device to a temperature change is bottom-up or stimulus driven. We argue that high level cognitive processes (as well as other factors such as experience) determine how the attentional control system is set, but that given that setting, the on-line response to events is purely stimulus driven, with no role played by high-level cognitive processes" (Folk et al., 1993, p. 682).

In line with this hypothesis, in the experiments reported here (see also Meyer et al., 1991; Niepel et al., 1994), attentional capture by schema-discrepant events depended on the relation between stimulus properties (the RV word) and subject characteristics (the previously acquired schema). More precisely, attention was captured by those properties of the RV word that exceeded the variable constraints of the currently activated schema. Presumably, this kind of attention capture is also purely stimulus driven, requiring no higher-level cognitive processes.

However, it should be noted that despite these basic common characteristics, attentional capture by schema-discrepant stimuli also differs in two important respects from the process of attentional capture observed in the experiments by Folk et al. (1992). In their experiments, which were conducted to test the contingent involuntary orienting hypothesis using a spatial cueing paradigm, a cue was presented for 50 ms at one of four possible target locations. After a 100 ms fixation display the target was presented for 50 ms. These authors found that the cue captured attention involuntarily only when it shared properties with the target and did not capture attention when it did not. Thus, the attentional control system was pre-set by properties of the target stimulus.

In the experiments reported here, in contrast, the attentional control system was set by the variable constraints of the currently active schema, which in turn were determined by the previous experience of the subjects. In addition, in the Folk et al. experiments, attention was captured by those properties of a cue which it shared with the target. In contrast, the attention-capturing effect of schema-discrepant stimuli is due to properties which they do not share with the currently activated schema.

In sum, then, the attentional system seems not only sensitive to goal-relevant stimulus properties, but also to schema-discrepant ones. As a matter of fact, the latter kind of attentional capture seems to correspond more closely to the thermostat analogy used by Folk et al. (1992) than the situation examined in their own experiments, because a thermostat responds only when the temperature deviates from the top-down determined set-point.

Both mechanisms of attention capture, however, are important for successful person-environment transactions: Attention capture by goal-relevant events permits the efficient preparatory allocation of processing resources in ongoing person-environment transactions, whereas attention capture by schema-discrepant events is a mechanism that enables, by means of schema revision or knowledge-updating, an effective prediction and control of future occurrences of the previously schema-discrepant event.

Alternative accounts of the present findings

In the concluding paragraphs I would like to briefly discuss two possible alternative accounts of the present findings. The first alternative account, that was already briefly considered with respect to the RT findings in Experiments 1 and 2, concerns neural network models. It was suggested there that although simple neural network models may be able to explain the RT increases in the first critical RV trial in the experimental groups with varying numbers of uniform NV trials preceding the critical one, they are at variance with the RTs in the subsequent RV trials. Already in the first RV trial following the critical one RTs were no longer delayed, suggesting an abrupt learning process. This finding poses a problem for simple neural network models because in these models learning is always a slow and continuous rather than an abrupt process (due to the assumption of a constant and low learning rate).

However, it appears that this problem is not an inherent limitation of neural network models in general. Recently, Phaf, Mul and Wolters (1994) proposed a novel neural network architecture that allows networks to learn quickly by means of an abrupt increase in the learning rate. More precisely, these authors assume that familiar stimuli cause only minimal competition among the network units and that as a consequence, only base-rate learning takes place. In contrast, novel stimuli (i.e., stimuli that are not yet represented in the network), are assumed to produce much competition among the network units. This, in turn, is assumed to cause a sudden increase in the learning rate and to instigate a random exploratory activation process that helps to resolve the competition and to form a new representation. Hence, this model seems to be capable of explaining the present RT findings if one assumes that the first RV presentation causes competition among the network units which, in turn, results in an abrupt increase in learning rate and the formation of a new representation.

Even if one grants that sophisticated neural network models such as the one just described are able to handle the RT findings, the question remains if and how such a model would account for the findings concerning the two remaining surprise indicators measured in the present experiments. Phaf et al. (1994) do not specify, for example, the relation between the amount of competition among the network units caused by the novel stimulus, the time needed to form a new representation and the memory performance for this new representation. Therefore, no precise prediction concerning memory performance for the RV word in the experimental groups can be derived from this model.

Furthermore, although the model is explicitly concerned with the consequences of novel stimuli within neural networks, this does not imply that it is also able to handle the consequences of (not necessarily new) surprising events. No surprise or similar mechanism is incorporated in this model. As a consequence, it remains mute with respect to the subjective experience of surprise investigated in the present experiments and other surprise components such as the facial expressions (cf. Meyer et al., 1994). In conclusion, it appears that existing neural network models do not offer a complete account of the present findings and therefore are not, at present, fully satisfactory alternatives to the proposed schema-theoretical approach. However, this does of course not exclude the possibility that more satisfactory neural networks could be developed in the future.

Finally, I would like to briefly consider a second alternative account of the present findings that was proposed by a reviewer. To explain some cases of surprise, Kahneman and Miller (1986; see also Kahneman, 1995) proposed "a supplement to the generally accepted idea that events in the stream of experience are interpreted and evaluated by consulting precomputed schemas" (p. 136).

These authors assume that surprise is sometimes produced by a contrast between an event and its counterfactual alternatives which are constructed after the occurrence of the event. Kahneman and Miller (1986) were mainly concerned with the elicitation of surprise rather than with its consequences. Therefore, their account seems mainly applicable to the subjective feeling of surprise. Presumably, this alternative account would hold that the intensity of surprise increases with the availability of counterfactual alternatives. After 32 NV trials, a counterfactual alternative to the RV presentation (namely, another NV trial) should be more readily available than after two NV trials; therefore, the feeling of surprise should be more intense in the 33-trials group.

To account for the present findings concerned with action delay and the focusing of attention on the surprise-eliciting event strictly in terms of counterfactual thinking seems less easy. In fact, one would predict effects contrary to those found. First, because after 32 NV trials a counterfactual alternative should be more readily available than after two NV trials, the interruption of ongoing processes should be shorter in the 33-trials group than in the 3-trials group. Second, because the construction of counterfactual alternatives is a backward-thinking process, the focus of attention is on relevant past events stored in memory. This focus of attention should, if anything, interfere with paying attention to the surprising event. Therefore, recall of the RV word (the index of the focusing of attention) might be expected to be reduced in the experimental groups. None of these predictions is supported by the present findings. Besides, counterfactual thinking per se cannot explain why the attention is focused on the surprising event after all.

However, it is possible that Kahneman and Miller's and the present schema theoretical approach to surprise differ only with respect to the processes that elicit surprise, whereas they are in agreement with respect to the components and the consequences of surprise. The main question would then be, whether or not the present experiments realized a case of the elicitation of surprise through the contrast between an event and a precomputed schema or a postcomputed counterfactual alternative. Kahneman and Miller (1986) themselves point out that the "notion of a schema is most useful in dealing with information that is needed often and that can be applied time after time" (p.139). These are precisely the conditions realized in the present experiments, thus backing up the appropriateness of the present schema-theoretical stance.

References

Belmore, S.M. (1987). Determinants of attention during impression formation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 480-489.

Bem, D.J. (1972). Self-perception theory. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 6, pp. 1-62). New York: Academic Press.

Berlyne, D.E. (1960). Conflict, arousal, and curiosity. New York: McGraw-Hill.

Bindra, D. (1959). Stimulus change, reactions to novelty, and response decrement. Psychological Review, 66, 96-103.

Bruce, D. & Gaines, M.T. (1976). Tests of an organizational hypothesis of isolation in free recall. Journal of Verbal Learning and Verbal Behavior, 15, 59-72.

Cermak, L.S. & Craik, F.I.M. (1979). Levels of processing in human memory. Hillsdale, NJ: Erlbaum.

Charlesworth, W.R. (1964). Instigation and maintenance of curiosity behavior as a function of surprise versus novel and familiar stimuli. Child Development, 35, 119-128.

Charlesworth, W.R. (1969). The role of surprise in cognitive development. In D. Elkind & J.H. Flavell (Eds.), Studies in cognitive development (pp. 257-314). New York & London: Oxford University Press.

Crocker, J., Fiske, S.T., & Taylor, S.E. (1984). Schematic bases of belief change. In R. Eiser (Ed.), Attitudinal judgment (pp. 197-226). New York: Springer.

Crutcher, R.J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. Psychological Science, 5, 241-244.

Darwin, C. (1872). The expression of the emotions in man and animals. London: John Murray.

Davidson, D. (1994). Recognition and recall of irrelevant and interruptive atypical actions in script-based stories. Journal of Memory and Language, 33, 757-775.

Desai, M.M. (1939). Surprise: A historical and experimental study. British Journal of Psychology, Monographs Supplement, 22, pp. 124.

Descartes, R. (1984). Die Leidenschaften der Seele (Passions of the soul). Hamburg: Meiner. (Original work published 1649: Les passions de l'âme.)

Dickinson, A. (1980). Contemporary animal learning theory. Cambridge: Cambridge University Press.

Ekman, P. & Friesen, W.V. (1975). Unmasking the face. A guide to recognizing emotions from facial clues. Englewood Cliffs, NJ: Prentice-Hall.

Ericsson, K.A. & Simon, H. (1993). Protocol analysis: Verbal reports as data (Rev. ed.). Cambridge, MA: MIT Press.

Fabiani, M. & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 224-240.

Fazio, R.H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M.S. Clark (Eds.), Review of personality and social psychology (Vol. 11: Research methods in personality and social psychology, pp. 74-97). Newbury Park, CA: Sage.

Folk, C.L., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent an attentional control settings. Journal of Experimental Psychology: Human Perception and Performance, 18, 1030-1044.

Folk, C.L., Remington, R.W., & Johnston, J.C. (1993). Contingent attentional capture. Journal of Experimental Psychology: Human Perception and Performance, 19, 682-685.

Frijda, N.H., Ortony, A., Sonnemans, J., & Clore, G.L. (1992). The complexity of intensity. Issues concerning the structure of emotion intensity. In M.S. Clark (Ed.), Review of personality and social psychology (Vol. 13: Emotion, pp. 60-89). Newbury Park, CA: Sage.

Graesser, A.C. (1981). Prose comprehension beyond the word. New York: Springer-Verlag.

Graesser, A.C., Gordon, S.E., & Sawyer, J.D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. Journal of Verbal Learning and Verbal Behavior, 18, 319-332.

Green, R.T. (1956). Surprise as a factor in the von Restorff effect. Journal of Experimental Psychology, 52, 340-344.

Green, R.T. (1958). Surprise, isolation, and structural change as factors affecting recall of a temporal series. British Journal of Psychology, 49, 21-30.

Hastie, R. (1984). Causes and effects of causal attribution. Journal of Personality and Social Psychology, 46, 44-56.

Holyoak, K.J., Koh, K.K., & Nisbett, R.E. (1989). A theory of conditioning: Inductive learning within rule-based default hierarchies. Psychological Review, 96, 315-340.

Hunt, R.R. & Mitchell, D.B. (1982). Independent effects of semantic and nonsemantic distinctiveness. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 81-87.

Isaacs, N. (1930). Children's "why" questions. In S. Isaacs (Ed.), Intellectual growth in young children (pp. 291-349). London: Routledge & Kegan Paul.

Kagan, J. (1971). Change and continuity in infancy. New York: Wiley.

Kahneman, D. (1995). Varieties of counterfactual thinking. In N.J. Roese & J.M. Olson (Eds.), What might have been: The social psychology of counterfactual thinking (pp. 375-396). Mahwah: NJ: Erlbaum.

Kahneman, D. & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. Psychological Review, 93, 136-153.

Kamin, L.J. (1969). Predictability, surprise, attention, and conditioning. In B.A. Campbell & R.M. Church (Eds.), Punishment and aversive behavior (pp. 279-296). New York: Appleton-Century-Crofts.

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

Mandler, J.M. (1984). Stories, scripts, and scenes: Aspects of schema theory. Hillsdale, NJ: Erlbaum.

McClelland, J.L. & D.E. Rumelhart. (1988). Explorations in parallel distributed processing. Cambridge, MA: MIT Press.

Metcalfe, J. (1995). A computational modeling approach to novelty monitoring, metacognition, and frontal lobe dysfunction. In J. Metcalfe & A.P. Shimamura (Eds.), Metacognition: Knowing about knowing (2nd printing, pp. 137-156). Cambridge, MA: MIT Press.

Meyer, W.-U. (1988). Die Rolle von Überraschung im Attributionsprozess [The role of surprise in the attribution process]. Psychologische Rundschau, 39, 136-147.

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., Niepel, M., & Schützwohl, A. (1994). Überraschung und Attribution [Surprise and attribution]. In F. Försterling & J. Stiensmeier-Pelster (Eds.), Attributionstheorie [Attribution theory] (pp. 105-121). Göttingen: Hogrefe.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1995). A process model of surprise. Unpublished manuscript. University of Bielefeld, Germany.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (in press). Towards a process analysis of emotions: The case of surprise. Motivation and Emotion.

Nelson, T.O. (1996). Consciousness and metacognition. American Psychologist, 51, 102-116.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A.F. Sanders (Eds.), Cognition and motor processes (pp. 255-293). Berlin: Springer.

Niepel, M., Rudolph U., Schützwohl, A., & Meyer, W.-U. (1994). Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. Cognition and Emotion, 8, 433-452.

Öhman, A. (1979). The orienting response, attention, and learning: An information-processing perspective. In H.D. Kimmel, E.H. van Olst, & J.F. Orlebeke (Eds.), The orienting reflex in humans (pp. 443-472). Hillsdale, NJ: Erlbaum.

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

Packer, J.S. & Siddle, D.A.T. (1989). Stimulus miscuing, electrodermal activity, and the allocation of processing resources. Psychophysiology, 26, 192-200.

Pearce, J.M. & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review, 87, 532-552.

Phaf, R.H., Mul, N.M., & Wolters, G. (1994). A connectionist view on dissociations. In C. Umilta & M Moscovitch (Eds.), Attention and perfomance 15: Conscious and nonconscious information processing (pp. 725-751). Cambridge, MA. MIT Press.

Posner, M.I. & Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Rescorla, R.A. (1988). Pavlovian conditioning: It's not what you think it is. American Psychologist, 43, 151-160.

Rescorla, R.A. & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), Classical conditioning II (pp. 64-99). New York: Appleton-Century-Crofts.

Ritter, D. (1997). Zur Entstehung von Intensitätsurteilen von Emotionen am Beispiel Überraschung. [On the construction of intensity judgments of emotion: The case of surprise]. Unpublished thesis. University of Bielefeld, Germany

Rumelhart, D.E. (1984). Schemata and the cognitive system. In R.S. Wyer & T.K. Srull (Eds.), Handbook of social cognition, (Vol. 1, pp. 161-188). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E. & Norman, D.A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In J.W. Cotton & R. Klatzky (Eds.), Semantic factors in cognition (pp. 37-53). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E. & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), Schooling and the acquisition of knowledge (pp. 99-135). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E., Smolensky, P., McClelland, J.L., & Hinton, G.E. (1986). Schemata and sequential thought processes in PDP models. In J.L. McClelland & D.E. Rumelhart (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models (pp. 7-57). Cambridge, MA: MIT Press.

Schmidt, S.R. (1991). Can we have a distinctive theory of memory? Memory and Cognition, 19, 523-542.

Schützwohl, A. (1997a). The structure of the intensity of surprise. Manuscript submitted for publication.

Schützwohl, A. (1997b). Responses to affectively neutral, positive, and negative schema-discrepant events. Manuscript in preparation. University of Bielefeld, Germany.

Schützwohl, A. & Horstmann, G. (in press). Überraschung, Handlungsunterbrechung und Schemarevision [Surprise, action delay, and schema revision]. In M. Jerusalem & R. Pekrun (Eds.), Emotion, Motivation und Leistung [Emotion, motivation, and performance]. Göttingen: Hogrefe.

Schützwohl, A. & Reisenzein, R. (in press). Children's and adults' reactions to a schema-discrepant event: A developmental analysis of surprise. International Journal of Behavioral Development.

Shand, A.F. (1914). The foundations of character. London: Macmillan.

Siddle, D.A.T. (1991). Orienting, habituation, and resource allocation: An associative analysis. Psychophysiology, 28, 245-259.

Siddle, D.A.T. & Packer, J.S. (1987). Stimulus omission and dishabituation of the electrodermal orienting response: The allocation of processing resources. Psychophysiology, 24, 181-190.

Sokolov, E.N. (1990). The orienting response, and future directions of its development. Pavlovian Journal of Biological Science, 25, 142-150.

Stiensmeier-Pelster, J., Martini, A., & Reisenzein, R. (1995). The role of surprise in the attribution process. Cognition and Emotion, 9, 5-31.

Thorndyke, P.W. & Hayes-Roth, B. (1979). The use of schemata in the acquisition and transfer of knowledge. Cognitive Psychology, 11, 82-106.

Tomkins, S.S. (1962). Affect, imagery, and consciousness. Vol. 1: The positive affects. New York: Springer.

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [On the effect of spheres formation in the trace field]. Psychologische Forschung, 18, 299-342.

Wagner, A.R. (1978). Expectancies and the priming of STM. In S.H. Hulse, H. Fowler, & W.K. Honig (Eds.), Cognitive processes in animal behavior (pp. 177-209). Hillsdale, NJ: Erlbaum.

Wallace, W.P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. Psychological Bulletin, 65, 410-424.

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

Wilson, T.D. (1994). The proper protocol: Validity and completeness of verbal reports. Psychological Science, 5, 249-252.

Author's notes

Achim Schützwohl, Department of Psychology, University of Bielefeld, Bielefeld, Germany.

I thank Marcus Ludewig who was responsible for the computer programs and Sabine Hollstegge and Veronika Koch who carefully ran the experiments. Rainer Reisenzein, Gernot Horstmann, Wulf-Uwe Meyer, and the reviewers made very helpful comments on earlier versions of the manuscript.

Correspondence concerning this article should be addressed to Achim Schützwohl, Universität Bielefeld, Abteilung Psychologie, Postfach 100 131, 33501 Bielefeld, Germany

Tables

Table 1. Percentage of subjects in the experimental and control groups with 3, 13, 23, or 33 trials, who referred to the RV word in the critical trial as an unexpected event, and recalled the RV and NV word and the dot's position in the critical trial, as well as their mean surprise ratings.

Table 2. Questions contained in the two different questionnaires used in Experiment 4.

A. Feeling of surprise

(1) How surprised were you about the black bar in the last trial? (*overall surprise intensity*)

(2) How intense was the feeling of surprise at its peak? (*peak amplitude*)

(3) How long did the feeling of surprise last? (*duration*)

B. Analysis of the surprising event and schema revision

(1) I made sure that one word was indeed presented inside a black bar. (*verification*)

(2) I wondered whether the presentation of the black bar has something to do with my task. (*action relevance*)

(3) I wondered why one word was presented inside a black bar in the last trial. (*causal analysis*)

(4) I was thinking about how to evaluate the occurrence of the black bar (something positive, something negative, or neutral). (*well-being*)

(5) I adjusted myself for further presentations of the black bar in upcoming trials. (*schema revision*)

Table 3. Mean ratings for the various aspects of the feeling of surprise and for the duration of the event analysis and schema revision processes in the 3-trials and the 33-trials groups in Experiment 4.

Figure Captions

Figure 1. The sequence of the presentation of the stimuli for a given trial, and examples of the words presented in NV trials and in the RV trial.

Figure 2. Mean differences between the RTs in the critical trial and the Baseline for experimental and control groups in Experiment 1 with 3, 13, 23, or 33 trials, respectively.

Figure 3. Mean differences (upper panel) and mean percentage differences (lower panel) between the RTs in the critical trial and the Baseline for experimental groups in Experiment 2 with 3, 13, 23, or 33 trials, respectively.

Figure 4. Mean differences between the RTs in the critical trial and the Baseline for experimental and control groups with constant and varied font-types in Experiment 3.

Footnotes

1. Notice that the present model shares many assumptions regarding the consequences of schema-discrepant events with G. Mandler's theory of cognition. The major difference between the two approaches consists of their scope: Whereas Mandler's theory is intended as a theory of emotion elicitation in general, the present analysis is restricted to the emotion of surprise.

2. The range of this scale was based on the range of RT increases usually obtained in the critical trial of the present experiments.

3. The three-way between-subjects ANOVA of the percentage RT increases revealed the same significant main effect for the number of trials.