Consciousness of Attention and Expectancy as Reflected in Event-Related Potentials and Reaction Times

Werner Sommer, Juliana Matt, and Hartmut Leuthold
Universität Konstanz
Konstanz, Federal Republic of Germany

The effects of conscious expectancies and attention on event-related potentials (ERP) and choice reaction times (RT) and their modulation by stimulus sequence were studied. Subjects retrospectively reported their expectancy of, and attention to, the terminal tones of short series. ERPs and RTs showed the usual sequential effects that were modulated by practice. As ratings were affected by only a few of the stimulus sequence, conscious access to sequence-based expectancy or attention appears to be fragmentary. Increased P300 amplitude with attention indicates conscious access to processing capacity. RTs and P300 latencies suggest stimulus processing time to decrease with sequence-based and consciously accessible expectancy. Differential effects of stimulus sequences and conscious expectancies on P300 amplitude indicate influences of two varieties of expectancy.

Expectancy is an important concept both in theories of the P300 component in event-related potentials (ERP) and for the explanation of the effects of stimulus sequences on reaction times (RT). The P300 is an electrically positive deflection in the ERP at 300 or more milliseconds, elicited by task-relevant stimuli. A model of the variables controlling the amplitude of this component has been presented by Johnson (1988). According to this model,

\[ \text{P300 Amplitude} = f[T \times (1/P + M)] \]

where \( T \) represents the amount of information transmitted by a stimulus, \( P \) reflects its subjective probability, and \( M \) corresponds to the meaning or significance of a stimulus.

Together with the prior probability of a stimulus class, expectancy determines the subjective probability variable in Johnson’s model and has most often been studied in the context of sequential effects. In the first systematic study of sequential effects on ERPs (Squires, Wickens, Squires, & Donchin, 1976), one of two tones of different pitch, presented in random order (Bernoulli series), had to be counted. The amplitude of the ERP elicited by a given stimulus was found to be inversely proportional to expectancy, which was determined according to the following model:

\[ \text{Expectancy} = M + A + P + \text{constant} \]

where \( P \) stands for the prior probability of the stimulus and \( M \) represents the exponentially decaying contribution of prior stimuli at specific sequential positions to the expectancy for the presentation of a like stimulus. In other words, the more recently and the more often a particular class of stimuli has been presented, the stronger the expectancy for this kind of stimulus to be presented at the next trial and the smaller the ERP to this kind of stimulus if it is actually presented. This basic expectancy for stimulus repetitions is modulated by an expectancy for the continuation of alternation patterns \( (A) \) in the stimulus sequence. Sequential effects in ERPs in accord with this model have been reported many times for counting tasks (Duncan-Johnson & Donchin, 1977; Johnson & Donchin, 1980, 1982; Squires, 1977; Petuchowski, Wickens, & Donchin 1977), prediction tasks (Chesney & Donchin, 1979; Munson, Ruchkin, Ritter, Sutton, & Squires, 1984; Tueting, Sutton, & Zubin, 1970) and choice reaction tasks (Duncan-Johnson, Roth, & Kopell, 1984; Ford, Duncan-Johnson, Pfef-ferbaum, & Kopell, 1982).

In choice reaction times, two kinds of sequential effects are commonly observed when different stimuli are presented in Bernoulli series and when response stimulus intervals (RSIs) are longer than 500 ms. First, the RT to a stimulus is shorter when it is preceded by a different stimulus than when it is preceded by a like stimulus (Hale, 1967; Kirby, 1972, 1976; Moss, Engel, & Faberman, 1967; Soetens, Boer, & Hueting, 1985; Williams, 1966; for a review, see Kirby, 1980). This so-called first-order alternation effect is attributed to subjects’ tendency to expect an excess of stimulus alternations over repetitions in random stimulus series. Note the contrast to the notion of subjects expecting stimulus repetitions in the model of Squires and coworkers. Occasional reports of first-order repetition effects at long RSIs are attributed to the use of few trials (Schvaneveldt & Chase, 1969) or to a complex relationship between stimulus and response (Bertelson & Ren-kin, 1966; Entus & Bindra, 1970).
A second kind of sequential effect is observed after continued runs of stimulus repetitions or stimulus alternations, where RT decreases with the length of that run. Conversely, RT increases with run length if the run is discontinued by the current stimulus (Remington, 1969; Soetens et al., 1985). These so-called higher order sequential effects are attributed to the consequences of expecting the continuation of a run of events. In either case, a stimulus that matches the expectancy is considered to be processed faster than a stimulus that does not.

There are discrepant views regarding the expectancy-generating processes as automatic or as controlled and whether or not these expectancies are conscious (i.e., subjectively accessible). In reaction time research with long RSIs, on the one hand, the buildup of sequence-based expectancies is considered a slow, controlled process (Kirby, 1980; Soetens et al., 1985), at the end of which there should be a consciously accessible expectation concerning the forthcoming stimulus (Shiffrin & Schneider, 1977). In ERP research, on the other hand, the expectancies underlying the sequence-related modulation of P300 are usually considered to be based on automatic processes (Johnson, 1988), and there is a tendency to view these expectancies as largely unconscious (Johnson & Donchin, 1982; Karis, Chesney, & Donchin, 1983). Even this view, however, is not entirely unitary. In a recent controversy (Donchin & Coles, 1988; Verleger, 1988), Verleger claimed that expected rather than unexpected events elicit large P300 amplitudes. After long runs of repetitions, for example, subjects are supposed to "wait" an alternate, and when it occurs, a large P300 is elicited because of the closure of the perceptual epoch. This awaiting of particular stimuli on the basis of the stimulus sequence is explicitly understood as a conscious expectation.

When consciously accessible expectancies have been directly assessed by means of predictions, the subjects usually have predicted the repetition of the immediately preceding stimulus (Geller & Pitz, 1970; Hale, 1967; Jarvik, 1951; Munson et al., 1984; Schwanenfeld & Chase, 1969; but for an exception see Keeler, 1969, Experiment 3). After runs of stimulus repetitions, this so-called positive recency effect turns into a negative recency effect or "gambler's fallacy", that is, an increasing expectancy for a stimulus change (Hale, 1967; Jarvik, 1951). Particularly the gambler's fallacy is in clear contradiction to the model of Squires et al. (1976), which holds that expectancy increases with the length of a repetition run. Donchin (1981) and Donchin and Isreal (1980) have pointed to various heuristics that may interact with perceived probability and therefore bias subjects' predictions. They suggested that P300 amplitude may serve as a more reliable indicator of subjective probability than the overt prediction behavior of the subjects.

None of the prior studies has attempted to integrate RTs, ERPs, and subjective expectancy measures in relation to a complete mapping of stimulus sequences. In the present study, we presented short Bernoulli series of stimuli that were just long enough to induce sequential effects. Each tone required choice reactions, and at the end of each series self-reports about the expectedness of the terminal stimulus and about the amount of attention payed to this stimulus were requested. This allowed us to compare and evaluate, for each stimulus sequence, subjects' subjective measures of expectancy and attention, their reaction times, and measures of P300 amplitude and P300 latency.

Collecting self-reports after the stimulus in question represents an alternative approach to the usual prediction measures of expectancy. Such predictions (a) have been shown to interact with the processing of the forthcoming stimulus (Geller, 1975; Näätänen & Merisalo, 1977), (b) possibly interfere with the generation of the expectancies, and (c) are subject to heuristics and strategies of the subject (Kahneman & Tversky, 1974). Although retrospective measures may have disadvantages of their own, there is no alternative when one wishes to control the interstimulus interval, which has been shown to be of crucial importance for sequential effects (Kirby, 1980). By recording self-report measures not only across but also within the various stimulus sequences, we were able to assess both sequence-generated and sequence-independent expectancy effects and their interactions.

In addition to the expectancy ratings, self-reports of attention were also requested. Ignored stimuli have very rarely elicited a P300 (Duncan-Johnson & Donchin, 1977), and increasing perceptual load on a primary task reduces the P300 to target stimuli of a concurrent secondary task (Donchin, Kramer, & Wickens, 1986). Accordingly, it is of interest whether P300 amplitude varies also as a function of self-reported attention. In Johnson's model of P300 amplitude, attention is one of the variables (7) determining the amount of information transmitted by the stimulus; correlations between self-reports of attention and ERPs would thus indicate conscious access to this variable.

The concurrent recording of behavioral and ERP variables can be particularly informative when these variables can be dissociated by some kind of experimental manipulation (Donchin, 1984). The considerable practice provided by the extensive data collection in the present study represents such a manipulation. Practice has been found not only to diminish RTs in general (Leonard, 1958; Mowbray & Rhodees, 1959; see also Teichner & Krebs, 1974), but also to reduce their modulation by the preceding stimulus sequences (Soetens et al., 1985; Vervaeck & Boer, 1980); however, there are virtually no ERP data available on the relation of practice to sequential effects. It would be interesting to know whether the sequential effects observed in RTs and ERPs, which seem so neatly associated at first sight, would remain associated over extended practice or become dissociated. Information about the relative practice effects might bear on the question of where in the information flow sequence-based expectancies act, and how closely coupled the sequential effects in ERPs and RTs actually are.

Method

Subjects

Complete sets of data were obtained from six women and four men of mean age 25.2 and 24.8 years, respectively. Two subjects had to be excluded after Session 1 because of excessive errors (approximately 60% each) or electrooculogram (EOG) artifacts. The 10 re-
remaining subjects were strongly right-handed, with laterality quotients greater than +66 (Oldfield, 1971). In order to maximize practice effects, we included only subjects who did not have prior experience in RT or ERP studies.

**Stimuli and Apparatus**

Auditory stimuli were two equiprobable sinusoidal tones of 1,000 and 1,500 Hz, both of 78.0 dB(A) intensity and 65-ms duration, including 8-ms rise/fall times, presented binaurally in random order through Sony MDR M66 headphones. A room ventilator provided a masking noise of 42 dB(A). The interstimulus interval was 1 s. During the presentation of the tone series a white fixation spot 26 mm in diameter was constantly presented on the monitor of a Commodore Amiga 2000 microcomputer at a distance of 1 m. Two telegraph keys mounted 20 cm apart on a horizontal response panel in front of the subject were used for recording RTs to the tones. The index fingers of the left and right hand operated the left and right key, respectively. The subjects were instructed to press one of the keys in response to the low tone and the other key to the high tone. The relation between tone pitch and response key was counterbalanced across subjects.

**Procedure**

The tone series was frequently interrupted after a number of tones to obtain subjects' self-reports. Interruptions occurred after no less than five to seven tones (randomly determined), only after at least five consecutive correct reactions, and only if the ERP to the terminal tone had been free of ocular artifacts. A reaction was accepted as correct if the proper key had been pressed and if the RT was between 100 and 800 ms. ERPs were accepted as artifact free if the EOG excursions during the recording epoch were less than 50 μV.

One second after a tone that met these criteria had been registered, self-reports were prompted by the simultaneous appearance of three vertical columns on the Amiga monitor. Self-reports were made with a mouse-driven cursor by selecting one of the columns. The columns were equally spaced horizontally, and each was approximately one fifth of the screen's width; the height of the left column was one third that of the screen's height, the middle column was two thirds of the screen's height, and the height of the right column was equal to that of the screen. The kind of self-report the subject was to make was indicated by the color of the columns. If they were blue, their heights corresponded to increasing degrees of attention, which the subjects had been able to pay to the terminal tone preceding the interruption. If the columns were red, their heights corresponded to different expectancies concerning this terminal tone. The large column was to be chosen when the terminal tone had been expected; the small one when the alternative, not presented tone had been expected; and the intermediate column when neither tone had been expected. The columns remained on the screen until the rating had been completed; a minimum time of 2 s was used to discourage hasty reporting. Immediately upon completion of the first rating, the second prompt was presented. The order of attention and expectancy prompts was altered. After each report the cursor was homed to the upper left-hand corner of the screen, and the first tone of the next series was presented 1 s after the second report.

Subjects were instructed in writing as to the experimental procedure and advised to avoid muscular and eye movements or blinks during the experiment proper. In addition to a base pay of DM 7.50 per hour, correct reactions and short RTs were rewarded with 0.50 DM for each percentage point below a 10% error rate and for each unit of 10 ms below an average RT of 320 ms.

Each subject was tested in four sessions, 2 to 3 days apart, usually at the same time of day. In each session 416 tone series of variable, subject-controlled length were presented, with short breaks after each block of 104 series. The first 16 series of a session were considered warm-up trials and discarded from data analysis. The average number of tones presented were, in order of session, 3,264 (range: 2,762–4,055), 3,159 (2,746–3,715), 3,229 (2,800–3,992), and 3,566 (2,771–7,162). The increase of the range during Session 4 was caused by a technical problem for one subject that increased the number of tones in the series but did not affect the data recorded to the terminal tone before the interruption. Without that subject the average number of trials in Session 4 would have been 3,166 (2,771–3,952). The average length of the tone series preceding the self-reports was, in order of session, 7.9, 7.6, 7.8, and 8.6 tones (7.7 when the subject with the technical problem is disregarded).

**Recording and Data Analysis**

For the purposes of data analysis only the RTs and ERPs to the terminal tones were considered, to which the self-reports also refer. In order to be able to distinguish between the two types of self-reports, as well as between different stimulus sequences of maximal length, the data were collapsed over tone pitch. The pitch of a particular tone was considered only inasmuch as it constituted a repetition (R) or an alternation (A) of the pitch of the preceding tone. This allowed for the differentiation of 16 stimulus sequences, defined by the terminal tone and all possible combinations of the four preceding tones. Collapsing over tone pitch seemed justified also by the fact that pitch did not affect RT, P300 area, or P300 latency (see Table 1).

For the analysis of the effects of the tone sequences on the dependent variables, we consistently used the model of Soetens et al. (1985). The 16 sequences were subdivided into those for which the terminal tone before the self-reports represented either a repetition or an alternation of the second-to-last tone (first-order sequence). The 8 higher order sequences for each first-order sequence were ordered from three consecutive stimulus repetitions (RRR) to three alternations (AAA; compare also Figure 1). When calculating linear regressions over these higher order sequences in the analyses of variance (ANOVAs) we do not wish to imply equal spacing between these sequences; rather, we are using the linear regressions merely to check for the existence of monotonic trends over the higher order sequences and to compare their direction and strength over different conditions.

The electroencephalogram (EEG) from Fz, Cz, and Pz, referenced to linked earlobes, and the vertical EOG were recorded with Grass ESH Ag/AgCl electrodes and Beckman electrode electrolyte paste. The EEG and EOG signals were amplified with time constants of 5.5 s and low-pass filters set to 40 Hz (~3 dB attenuation, 12 dB rolloff/octave). All signals were sampled at 100 Hz for 900 ms, starting 90 ms prior to stimulus onset, and stored on magnetic tape.

The averaged ERPs were digitally low-pass filtered at 4.9 and 8.8 Hz (~3 dB; Ruchkin & Glaser, 1978) for the rating and the practice analysis, respectively. Then the average activity during a 90-ms pre-stimulus baseline was subtracted from each waveshape and an area measure of P300 (270 to 370 ms) was taken. These voltage integrals over time were divided by the number of time points involved, resulting in a measure representing the average amplitude during the

<table>
<thead>
<tr>
<th>Tone pitch (Hz)</th>
<th>Reaction time (ms)</th>
<th>P300 area (μV)</th>
<th>P300 latency (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>270.4</td>
<td>8.2</td>
<td>302</td>
</tr>
<tr>
<td>1500</td>
<td>263.8</td>
<td>8.1</td>
<td>300</td>
</tr>
</tbody>
</table>
Component interval in μV. P300 latency was measured following the procedure of Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983). A negative (electrical positivity to numerical negativity in our data) 2 Hz sinus half-wave template was aligned with the averaged ERP waveforms from the Pz electrode. The center of the template was moved across the ERP from 270 ms after stimulus onset to 370 ms in steps of 10 ms, and cross-products between template and ERP were calculated at each lag. At the lag of the maximum cross-product, P300 latency was taken as the time point in the ERP corresponding to the center of the template.

All dependent variables were subjected to repeated measures ANOVAs with the following factors: first-order stimulus sequence (R, A) and higher order sequence (RRR, ARR, RAR, AAR, RRA, ARA, RAA, AAA). The electrode sites (Fz, Cz, Pz) (for the P300 amplitude measures) and the sessions (1 to 4) (for the practice-related analysis) were also included as repeated measures factors in the ANOVA. If any main effect or interaction involving the higher order sequence was significant, the linear trends across the levels of this factor were also considered. In order to correct for possible violations of sphericity assumptions, conservative (Huynh–Feldt) F tests were used throughout, with degrees of freedom before and p values after the adjustment reported. For a posteriori comparisons, the levels of significance were corrected according to the Bonferroni criterion.

Sequential Effects in ERPs and RTs and Their Modulation by Practice

Results

Reaction times. Throughout all four sessions RTs were about 10 ms longer after first-order stimulus repetitions than after first-order alternations (see Table 2), F(1, 9) = 5.73, MS_e = 3,129.90, p < .05. If the terminal tone discontinued runs of higher order repetitions or alternations, RT was delayed in proportion to the length of that run (Figure 1), F(7, 63) = 35.47, p = .62, MS_e = 323, 309, 305, and 302 (M5 sessions 1 to 4). This delay diminished over the session, F(21, 189) = 154.22, p < .001. During Session 1 the P300 latency to first-order repetitions was smaller than the P300 to alternations (Table 2), F(1, 19) = 7.23, MS_e = 1,856.61, p < .05, but in the course of the remaining sessions the first-order effects gradually vanished, F(2, 18) = 1.04, p = .92, MS_e = 989.04.

Practice affected the differences in P300 amplitude between first-order repetitions and alternations, F(3, 27) = 5.74, p = 1.13, MS_e = 2,233.12, p < .01. During Session 1 the P300 of first-order repetitions was smaller than the P300 to alternations (Table 2). F(1, 19) = 7.23, MS_e = 1,856.61, p < .05, but in the course of the remaining sessions the first-order effects gradually vanished, F(2, 18) = 1.04, p = .92, MS_e = 989.04.

Event-related potentials. Figure 2 depicts sequential effects on the waveforms of the averaged ERPs from the Pz electrode. In Figure 3 first-order repetitions and alternations are highlighted by collapsing over higher order sequences, and the scalp topography is visualized by depicting all three leads. Like RT, the P300 latency (Figure 4) was increased when runs of repetitions or alternations were discontinued by the terminal stimulus, F(7, 63) = 16.9, p = .05, MS_e = 8.85, p < .001, and F(linear: 1, 9) = 43.0, MS_e = 154.22, p < .001. Although there was an overall shortening of P300 latency from Session 1 to Session 4 (M5 = 323, 309, 305, and 302 ms, respectively), F(3, 27) = 7.9, p = .12, MS_e = 17.81, p < .001, this shortening was independent of the stimulus sequence. P300 amplitude showed a centroparietal scalp topography, M5 (Fz to Pz) = 4.5, 10.3, and 9.3 μV, F(2, 18) = 20.6, p = .91, MS_e = 29,176.25, p < .001. Similar to the RT, it increased with the length of repetition and alternation runs, if discontinued by the terminal stimulus (Figure 5), F(7, 63) = 17.7, p = .66, MS_e = 2,668.64, p < .001, and F(linear: 1, 9) = 30.93, MS_e = 28,770.15, p < .001.

Table 2

<table>
<thead>
<tr>
<th>Session</th>
<th>R</th>
<th>A</th>
<th>R</th>
<th>A</th>
<th>R</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction time (ms)</td>
<td>280</td>
<td>273</td>
<td>324</td>
<td>323</td>
<td>6.2</td>
<td>8.2</td>
</tr>
<tr>
<td>P300 latency (ms)</td>
<td>272</td>
<td>258</td>
<td>313</td>
<td>305</td>
<td>7.8</td>
<td>8.0</td>
</tr>
<tr>
<td>P300 area (μV)</td>
<td>274</td>
<td>260</td>
<td>309</td>
<td>301</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>264</td>
<td>255</td>
<td>301</td>
<td>302</td>
<td>8.2</td>
<td>8.1</td>
<td></td>
</tr>
</tbody>
</table>

Note. R = stimulus repetition; A = stimulus alternation.

Figure 1. Sequential effects on choice reactions time (RT) of the four preceding stimuli in a random series of two equiprobable tones of different pitch. (RTs are collapsed over pitch; only repetitions (R) and alternations (A) are plotted. RTs to terminal tones are plotted separately for first-order repetitions (R, left side) and alternations (A, right side) of the preceding tone. The eight higher order sequences within each first-order sequence are ordered from three repetitions (RRR) to three alternations (AAA). The RTs are superimposed for Sessions 1 to 4.)
We will discuss higher order and first-order sequential effects in turn, first for Session 1 and then for the effects of practice in the course of the remaining four sessions. An overview of the findings is presented in Table 3.

**Higher order effects.** The higher order effects observed during Session 1 conform to those reported by others and have usually been interpreted to indicate that subjects expect runs of stimulus repetitions or stimulus alternations to continue. When those runs are actually discontinued, prolonged RTs (e.g., Remington, 1969; Soetens et al., 1985), increased P300 amplitudes (Chesney & Donchin, 1979; Duncan-Johnson & Donchin, 1977, 1982; Duncan-Johnson et al., 1984; Johnson & Donchin, 1980, 1982; Munson et al., 1984; Squires et al., 1976, 1977), and increased P300 latencies (Duncan-Johnson et al., 1984; Ford, Pfefferbaum, & Kopell, 1982) have been considered as indicative of surprise.

As to the consequences of practice for the higher order sequential effects on RTs, the study most directly comparable to ours is Experiment 3 of Soetens et al. (1985). At RSIs of 500 ms the RTs initially showed the usual higher order effects and a first-order alternation effect. With practice, both first-order and higher order effects diminished. Although they used very short RSIs (160 ms), Vervaeck and Boer (1980) found expectancy effects after long runs of alternations, facilitating when continued and inhibiting when discontinued. After 1,000 trials of practice this expectancy effect disappeared. Both Soetens et al. and Vervaeck and Boer interpreted this practice-dependent reduction of sequential effects as a consequence of a global decrement of expectancies.

Although practice in the present study decreased the higher order effects for RTs, it did not affect those of P300 amplitude or latency. These data do not conform with the hypothesis of a global decrement of the expectancies generated by the preceding sequence, because the costs of expectancy did not decrease (shortening of the RTs to discontinued stimulus runs) and the benefits of expectancy in continued runs did not diminish. These sequences continued to produce the shortest RTs throughout all four sessions. The robustness of the expectancies against practice is also indicated by the stability of the higher order sequential effects in P300 amplitude.

The locus of the practice effects on the RTs may be brought into better focus by considering P300 latency. P300 latency is usually considered to reflect the time it takes to evaluate and categorize an event and has been shown to be largely independent of response selection and execution processes. When speed rather than accuracy is emphasized, the response may precede the P300; in this case, it is argued, the response is based on only a partial analysis of the stimulus, whereas P300 latency includes the time of evaluation. The existence of higher order effects on P300 latency demonstrates that stimulus evaluation and categorization processes are influenced by expectancy. The stability of these effects over sessions indicates that this influence does not change with practice. In addition to the sequential effects on stimulus evaluation, there seems to be an additional effect on processes following stimulus evaluation (e.g., response selection or execution processes) because, first, the sequential effects in RTs are considerably stronger than in P300 latency (cf. Figures 1 and 4) and, second, only the sequential effects in the RTs are affected by practice. Therefore, sequence-based expectancies seem to act upon at least two stages of the information flow, and only the expectancy effects on the later stages are diminished by practice, whereas those on early stages are stable. On the basis of the present data it cannot be decided, however, whether we are dealing with one expectancy-generating process, whose effects are selectively diminished, or whether there are two such processes, one that acts on stimulus evaluation and categorization and is robust.
CONSCIOUSNESS OF ATTENTION AND EXPECTANCY

against practice and one that acts on later processing stages and is susceptible to practice.

It should be noted here that there was a marked decrease of P300 latency over sessions independent of any stimulus sequences, which was accompanied by a similar, albeit insignificant, decrease of RT. This indicates that the sequence-independent reaction time shortening with practice might be explained by improvements in stimulus evaluation and categorization processes.

First-order effects. The first-order effects for P300 amplitude and RT during Session 1 agree well with the literature, but they are dissociated from each other. The RTs show a first-order alternation effect (i.e., shorter RTs to stimulus alternations than to stimulus repetitions). The alternation effect is consistent with most RT studies using RSIs of more than 500 ms. Contrary to the alternation effect in the RTs but in accord with virtually all ERP studies (Chesney & Donchin, 1979; Duncan-Johnson & Donchin, 1977, 1982; Duncan-Johnson et al., 1984; Johnson & Donchin, 1980, 1982; Munson et al., 1984; Squires et al., 1976, 1977), P300 amplitudes are smaller after a first-order repetition, that is, a first-order repetition effect.

One might seek the explanation for the conflicting first-order effects in procedural differences of the studies; the present study, however, shows a “first-order dissociation” between RTs and ERPs for the same data set. On the other hand, the two previous studies that have simultaneously assessed sequential effects in both RTs and ERPs (Ford, Duncan-Johnson, Pfefferbaum, & Kopell, 1982; Duncan-Johnson et al., 1984) showed conforming higher order as well as first-order (repetition) effects for both variables.

The discrepancy between the first-order effects in RTs and ERPs is further emphasized by the effect of practice, which, in the present experiment, was restricted to the ERPs. Although practice tended to diminish the discrepancy between the first-order effects of RTs and P300 amplitude after the first session, it did so by increasing the P300 to repetitions and thus abolishing the well-documented first-order repetition effect in P300.

As the higher order sequential effects both for RTs and P300 amplitudes are in accord with each other, both effects can be explained by expectancy. For the conflicting first-order effects, however, an alternative or at least additional explanation apart from expectancy must be found either for the RTs or the P300 amplitudes, because subjects cannot be considered to expect stimuli both to repeat and alternate. One clue may be provided by our rating data (see below). Although the relationship between the expectancy ratings and the stimulus sequences was generally weak, subjects (insignificantly) tended to rate stimulus alternations as more expected than repetitions. Therefore, we are inclined to assume a process in addition to expectancy for the first-order effects in P300 amplitude rather than for the RTs. A process is needed that acts specifically on repetitions and causes P300 to increase with practice. Thus, one should not assume, for example, that stimulus changes by themselves exert a P300-enhancing and practice-dependent influence, because it is the influence of stimulus repetitions that was altered with practice.

We tentatively suggest that for unpracticed subjects there may be more postresponse uncertainty as to the actual pitch of a tone if it has not been preceded by a different tone to which it can be contrasted. Similar to the reduction of P300

Figure 3. ERPs superimposed for first-order stimulus repetitions (solid lines) and alternations (broken lines) for Sessions 1 to 4 and for all three electrode sites.
amplitude in conditions of low discriminability and low decision confidence (Johnson, 1988), the relatively greater equivocation about stimulus identity after first-order repetitions might reduce P300 relative to first-order alternations. The increase of P300 amplitude to repeated stimuli with practice might then be related to the improvement of decision confidence. Similar increases of P300 with practice have been shown in learning experiments where P300 increased with decreasing equivocation about the correct response (Johnson, Pfefferbaum, & Kopell, 1985; Peters, Billinger, & Knott, 1977).

A conclusion of importance for the data yet to be described is that the experimental procedure used here produces sequential effects for RTs and ERPs that are very similar in their direction to those reported for continuous stimulus series (e.g., Soetens et al., 1985; Squires et al., 1976). Discrepancies emerged only in the context of practice and only for the ERPs where practice has not been studied before. Also, the rather short RTs and P300 latencies demonstrate that we have basically been able to retain the spontaneous nature of the task. Thus, we consider it likely that our findings can be generalized to other experimental situations generating sequential effects.

Figure 4. Sequential effects on P300 latency, superimposed for Sessions 1 to 4, measured at Pz as the time of maximum crosscorrelation of a 2 Hz sinus halfwave template between 270 and 370 ms after stimulus onset.

Figure 5. Sequential effects on P300 amplitude, superimposed for Sessions 1 to 4, measured at Pz as the average amplitude between 270 and 370 ms after stimulus onset, related to a 70-ms prestimulus baseline.

Sequential Dependencies in Self-Reported Expectancies and Attention

Results

In order to check for dependencies between the expectancy and attention ratings, a contingency table for the total frequencies of both ratings (over subjects, sessions, and sequences) was calculated (Table 4). Although the distribution of frequencies over cells was highly unequal ($\chi^2 = 254.8, p < .001$), the $\chi^2$-coefficient of .089 indicated that the association between the attention and expectancy ratings was negligible.

The expectancy ratings from each stimulus sequence and session were analyzed in two ways. First, the ratio of nonneutral expectancy ratings (when either the actually presented terminal tone $P$ or the alternative tone $nP$ had been expected) to all ratings given (including the cases when no particular expectancy had been held) was calculated. Second, a score for the relative expectedness of the terminal tone was calculated by $(P - nP)/(P + nP)$. The self-reports of attention were analyzed by scoring the ratings of increasing attention from 1 to 3 and averaging them for each sequence and session.
As shown in Figure 6, the stimulus sequences similarly affected both the self-reports of attention and the subjects' inclination to give nonneutral expectancies. With an increasing number of repetitions in the sequences, more nonneutral expectancy statements and higher attention ratings were given. This is reflected in main effects of the higher order sequence for the nonneutral expectancy statements and the attention ratings, \( F(7, 63) = 9.24 \) and \( 5.05, \) respectively, and in the linear trends over the higher order sequences: \( F(1, 9) = 9.95, p = .01, \) and \( F(1, 9) = 5.22, \) \( p = .04, \) respectively, and in the linear trends over the higher order sequences: \( F(1, 9) = 9.95, p = .01, \) and \( F(1, 9) = 5.22, \) \( p = .04, \) respectively. There were three sequences where the ratio of nonneutral expectancies was further increased—continued and interrupted runs of two repetitions (RRRR) and continued runs of three alternations (AAAA). This was reflected in an interaction of the first-order and the higher order effects, \( F(7, 63) = 3.32, p = .03, \) and \( F(3, 27) = 5.80, \) \( p = .01, \) respectively. A similar tendency for their subjects to predict the repetition of the terminal stimulus, averaged over sessions, was \( .20, \) after first-order alternations and \( .24, \) after first-order repetitions. The mean relative expectedness of the terminal stimulus, averaged over sessions, was \( .20 \) after first-order alternations and \( .24 \) after first-order repetitions. Similarly, there were no reliable effects of the first-order sequences on either of the two expectancy measures. The ratio of nonneutral expectancies was \( M = .49, \) after first-order repetitions and \( M = .42, \) after first-order alternations. The relative expectedness of the terminal stimulus was only \( M = .02, \) after first-order repetitions but \( M = .20, \) after first-order alternations. However, for both expectancy measures the first-order effects were not significant, \( F(1, 9) = 4.11, M_S = .19, \) \( p = .07, \) and \( F(1, 9) = 2.86, M_S = 1.94, p = .12, \) respectively.

In addition, the gambler's fallacy (Jarvik, 1951) was tested by analyzing the changes of the relative expectedness of the terminal stimulus with increasing numbers of preceding stimulus repetitions. The mean relative expectedness of the terminal stimulus, averaged over sessions, was \( .20, \) after first-order alternations (RRRR, ARRA, RARA, AARA, RAAR, ARAA, AAAAA, and AAAAA). .03 after one stimulus repetition (RRRR, ARRA, RARA, and AAAAA), and .04, .01, and .04 after two (RRRR and RARR), three (ARRR), and four repetitions (RRRR), respectively. Numerically, this conforms with the notion that increasing runs of repetitions lead to an increase of expectancy for an alternation, but statistical significance failed by a large margin (\( F < 1 \)).

**Discussion**

Systematic effects of the stimulus sequences comparable to those for RTs and ERPs were lacking for the self-reports of expectancy and attention. If the sequential effects on RTs or ERPs were related to or caused by systematic and consciously accessible variations in the kind of stimulus expected, sequential effects similar to those in RTs or ERPs should be observed in reports of the relative expectedness of the terminal tone as well. But this is evidently not the case, as demonstrated by the absence of any first-order effects, by the lack of a linear trend over the higher order sequences, and by the absence of a gambler's fallacy.

Distinct expectancies are exclusively confined to continued and discontinued runs of three alternations (AAAA, AAAR). This conforms with the findings of Munson et al. (1984) that, after runs of alternations, another alternation was usually predicted. On the other hand, Munson et al. found an overall tendency for their subjects to predict the repetition of the prior stimulus, whereas for our subjects first-order alternations were somewhat, but insignificantly, more expected than first-order alternations.
and in the ratio of nonneutral expectancies were observed

\[ \frac{nP}{P + nP} \]

expectancy.

Figure 6. Sequential effects on self-reports: the amount of attention paid to the terminal stimulus in a tone series; the ratio of nonneutral expectancies (expecting either the presented stimulus or its alter-
native to all ratings given, including neutral expectancies; and a measure of the relative expectedness of the presented stimulus \( \frac{P - nP}{P + nP} \).

In sum, there are stimulus sequences that do elicit heightened attention and readiness for nonneutral expectancy ratings, and there are also sequences that are related to particular directions of expectancy, but these findings cannot be directly related to the systematic sequential effects in RTs and ERPs. Thus, these data agree with the notion that sequence-generated expectancies do not have to be conscious (Johnson, 1988).

Self-Report–Related Effects on RTs and ERPs

Results

The frequency distributions of the attention and expectancy ratings were inevitably under the control of the subjects. Therefore, data frequencies were low for a number of cells, even after collapsing over sessions. For the attention-related data, 141 of the 480 cells (3 levels of ratings, 16 stimulus sequences, and 10 subjects) contained less than 10 ERPs or RTs and 73 contained only 5 or even fewer. For the expectancy-related data the respective figures are 75 and 44 cells, and 4 cells were empty. The empty cells were replaced by the mean RTs and ERP measures of the other subjects for this condition after digitally low-pass filtering the originally recorded ERPs at 4.9 Hz to improve the signal-to-noise ratio. Because the main effects and the interaction of first-order and higher order sequence, as well as the electrode site of the ANOVAs, merely restate results reported above, only effects of the rating categories and their interactions with the sequential effects or electrode sites will be reported below.

Attention ratings. There were no effects of increasing attention ratings on P300 latency, but the RTs tended to decrease about 10 ms (Figure 7), \( F(2, 18) = 3.32, \epsilon = .83 \), \( MS_e = 1,433.69, p = .07 \), and the P300 amplitude showed a highly significant increase of 1.5 \( \mu \)V (Figures 7, 8, and 9), \( F(2, 18) = 10.42, \epsilon = 1.10 \), \( MS_e = 3,115.01, p < .001 \). All these attention-related results were independent of the stimulus sequences.

Expectancy ratings. The analyses according to the three levels of the expectancy ratings revealed RTs that were shorter by about 25 ms to expected than to unexpected stimuli (Figure 10), \( F(2, 18) = 14.14, \epsilon = .91 \), \( MS_e = 1,440.37, p < .001 \). Similarly, P300 latency was shorter to expected than to unexpected stimuli by about 20 ms (Figures 10, 11, and 12), \( F(2, 18) = 7.17, MS_e = 19.57, \epsilon = .75, p < .01 \). Again, these effects were independent of the stimulus sequence.

In contrast, P300 amplitude showed an interaction between self-reported expectancy and the first-order stimulus sequence (Figures 10, 11, and 12), \( F(2, 18) = 6.18, \epsilon = .70 \), \( MS_e = 2,337.02, p < .05 \). After first-order stimulus repetitions, P300 was insignificantly larger by .5 \( \mu \)V when the terminal tone actually presented, rather than its alternative, had been expected: \( F(1, 9) = .72 \). After stimulus alternations, however, P300 was larger by 1.7 \( \mu \)V when the terminal tone rather than its alternative had been expected, \( F(1, 9) = 8.19, MS_e = 203,937.59, p < .05 \).
Discussion

The relatedness of expectancy and attention reports to differences in ERPs confirms and extends our earlier findings of subjective awareness of ERP differences (Sommer & Matt, in press). The dissociation of the expectancy effects on P300 amplitude and RTs makes it unlikely that one effect can be reduced to the other. Thus, the expectancy effects on P300 cannot be attributed to the differential contributions of key press–related motor potentials or to subjects basing expectancy reports on the discrimination of their own reaction speed. A similar argument pertains to a potential transfer between the attention and the expectancy ratings, which is unlikely because of the differential effects of these ratings on P300 amplitude. Inequality of cell frequencies is also not a likely source of the reported effects. Although cell frequencies were sometimes quite low at the highest order of factor combinations, as mentioned above, there were at least 40 and generally more observations per cell and subject even for the most complex effects discussed here. Furthermore, the outermost rating categories, low and high attention, and the two nonneutral expectancies were about equal in cell frequency distribution.

Additionally, it cannot be argued that the subjects' discrimination of ERPs and RTs by attention and expectancy ratings may be based on the sensory discrimination of the different stimulus sequences, for instance, by rating the presented terminal tone or its alternative as expected after continued or discontinued alternation runs, respectively. First, as reported above, the sequential effects on the expectancy ratings were weak. Second, even if subjects did base some of their ratings on the preceding stimulus sequences, the ERP and RT differ-

Figure 7. Effects of self-reported attention on P300 amplitude (left panel), reaction times, and P300 latency (right panel).

Figure 8. ERP waveshapes at Pz according to the attention ratings and all 16 stimulus sequences (cf. Figure 2).
The expectancies between the various ratings were independent of such strategies, because the differences are present within each sequence category (cf. Figure 11). This holds true also for the interaction of the first-order sequence and the expectancy ratings, which cannot be reduced to the discrimination of stimulus sequences.

The positive relationships between self-reported attention and P300 amplitude conform with findings from dual-task studies (Donchin, Kramer, & Wickens, 1986), indicating that P300 amplitude is related to the amount of processing resources allocated to a stimulus. According to our results, these variations of resource allocation may actually be introspectively accessible.

The expectancy-related findings in RTs and for the P300 latencies also conform well with the literature. The sequential effects in RTs for long RSIs are usually attributed to the expectedness or unexpectedness of the stimuli, diminishing or increasing the RTs (Kirby, 1980; Laming, 1969). Thus, it appears plausible, as well, that stimuli that are explicitly defined by subjects as unexpected are related to increased RTs, as seen in our data. These findings also conform to those of Experiment 3 of Williams (1966), where subjects had to predict which of two alternative stimuli would be presented next. After correct predictions, the RTs to the presented stimulus were shorter than after false predictions. Like in the present study, this effect was independent of the status of the stimulus in question as first-order repetition or alternation. The increase of P300 latencies after unexpected as compared with expected stimuli may reflect increased processing times for unexpected events (Squires, Hillyard, & Lindsay, 1973).

Much less easily accounted for are the findings for the P300 amplitude, which is somewhat (insignificantly) larger after unexpected as compared with expected stimuli for first-order stimulus repetitions but is markedly increased for expected events after stimulus alternations. This latter finding contradicts the common notion that unexpectedness is related to large P300 amplitudes. The evidence supporting this idea, reviewed, for example, by Johnson (1988), includes the sequential effects on P300 and the increase of P300 amplitude with the rareness of a stimulus.

Only a few studies, however, have directly related P300 amplitude to self-reports descriptive of expectancies about future events. Horst, Johnson, and Donchin (1980) used a paired associate learning task with nonsense syllables where, in addition to responding with an associate to the first syllable,
subjects had to rate their confidence about the correctness of the response. The amplitudes of the P300 elicited by the subsequently presented "response" syllable varied inversely with the subjective probability of the syllable. That is, large P300 amplitudes appeared when subjects were right but had expected to be wrong and when they were wrong but had expected to be right, whereas the confirmation of expectancies elicited small P300s.

Chesney and Donchin (1979; reported in more detail in Donchin, 1979, and Donchin & Isreal, 1980) had subjects predict the next stimulus in a visual Bernoulli series. The ERPs were averaged according to the first-order stimulus sequence, the prediction made, and whether the prediction was confirmed or disconfirmed (prediction outcome). The results for the P300 amplitude can be interpreted in two ways: (a) Subjects actually expect stimuli to repeat, and thus disconfirmations elicit larger P300 amplitudes than confirmation when repetitions are overtly predicted but smaller P300s when alternations are predicted; (b) more simply, alternations elicit larger P300s than repetitions, whatever the prediction.

In a similar prediction study in Bernoulli series of clicks and omissions, Munson et al. (1984) found a complex interaction among the first-order stimulus sequence, the subjects' predictions, and prediction outcome. Conforming with the quantitative trend in our data, a stimulus repetition elicited larger P300s when disconfirming a prediction. For stimulus alternations, however, the prediction was not related to P300 amplitude.

Whereas Munson et al.'s findings of larger P300 amplitudes to unexpected as compared with expected repetitions agree with ours, neither their results nor those of Chesney and Donchin show the most striking feature of our findings—the larger P300 amplitude to expected than to unexpected alternations. A possible explanation for the difference in findings between our study and these studies may be the difference in methods. Expectancies concerning particular events may be consciously accessible to a different degree when they are to be stated in advance of the event, as compared with the retrospective view taken in the present experiment.

Conclusion

The effects of conscious expectancies on P300 amplitude contrast with the effects of sequence-based effects in ERPs and RTs. The sequential effects conform with the literature and seem to be outcome dependent, revealing enhanced P300 amplitudes, longer P300 latencies, and increased RTs when runs of events are discontinued, that is, when subjects were presumably surprised by the (unexpected) terminal stimulus. According to the analysis of the expectancy ratings, such sequence-generated expectancies seem to be largely unconscious, as only 2 of the 16 stimulus sequences affected the ratings.

The comparatively short RTs and P300 latencies found here, as well as the poor conscious access to sequence-based expectancies, conform with the view that these expectancies are related to a fast, inflexible, automatic process rather than to a controlled process (Johnson, 1988; Johnson & Donchin,
1982; Karis et al., 1983). This is partially corroborated by our practice results, which show robust higher order sequential effects in ERP measures. That there are also practice-dependent changes in sequential effects in RTs is a fact that there might also be a controlled aspect in the workings of sequence-based expectancies. The relative inclination to conceptualize expectancies as products of controlled processes in RT work might be seen as related to these possible contributions of controlled processes to sequential effects in RTs but not in ERPs. But even if there is a controlled aspect to the sequential effects in RTs, it is not reflected in self-reports of expectancies, either because it is too weak or because it is unrelated to what is measured by the rating process.

When we asked subjects to report their expectancies, we may have been tapping those aspects of expectancy that are consciously accessible and that are related to P300 amplitude according to rules quite different from those governing the relatively inaccessible, direct sequential effects on P300. One interpretation of our different “expectancy” effects at the levels of stimulus sequences and of self-reports may follow Kahneman and Tversky’s (1982) suggestion that subjects may simultaneously hold different types of expectations, which may, in turn, be consciously accessible to different degrees. In contrast to the largely unconscious sequence-based expectancies, there do exist conscious expectancies, not explainable by differential contributions of the stimulus sequences and that would not have been predicted from traditional ERP research (e.g., Johnson, 1988; Squires et al., 1976). The finding that within first-order stimulus alternations consciously expected events are related not to particularly small but to large P300 amplitudes conforms more with Verleger’s view of greater P300 amplitudes to stimuli that have been “awaited.”

However, if the (insignificant) P300 findings for first-order repetitions (cf. Figure 10) are to be taken seriously, an even simpler interpretation of the effects of consciously accessible expectancies on P300 amplitude may be formulated. One might say that we found large P300 amplitudes whenever subjects had expected stimulus alternations and small P300 amplitudes when repetitions had been expected, with little relevance in either case as to whether the stimulus actually repeated or alternated. When formulated this way, the expectancy effects on P300 amplitude become completely independent of the actual first-order sequence and solely dependent on the sequence expected. This reformulated effect might be interpreted in at least two ways: (a) Any stimulus, be it a change or a repetition, is increased in relevance and thus elicits a comparatively large P300 (Donchin, Ritter, & McCallum, 1978) when environmental change rather than constancy is consciously expected; or (b) the internal representation of expectancy might be altered by the very processes that control the P300 amplitude—one may experience the expectation of a change of events when, by some chance, an event has elicited a large P300, and a repetition when P300 has been small.

In addition, our data indicate different relationships of consciously accessible expectancies and P300 amplitude, on the one hand, and RTs and P300 latencies, on the other. The latter are accelerated for expected events and suggest that not only sequence-based but also conscious expectancy can speed up processing time.

References


Cognition, 8, 476–488.


