Gender differences in late positive components evoked by human faces

JUAN C. OLIVER-RODRÍGUEZ, ZHIQIANG GUAN, AND VICTOR S. JOHNSTON

Abstract

Event-related potentials (ERPs) were recorded in male and female participants in response to 32 male and 32 female faces. Participants were instructed to simply look carefully at each face; after ERP collection they were asked to rate each face on a 5-point attractiveness scale. A positive correlation between average rating and average P300 scores to opposite sex faces was observed in male faces. Participants were instructed to simply look carefully at each face; after ERP collection they were asked to rate each face on a 5-point attractiveness scale. A positive correlation between average rating and average P300 scores to opposite sex faces was observed in male (r = .40) and in preovulatory (r = .41) and postovulatory (r = .44) female subjects. Correlations to same sex faces were only found in postovulatory females (r = .61). Male participants showed a much larger average P300 than did female participants, and the P300 evoked in female participants was unexpectedly larger to female than to male faces. Neither task relevance nor stimulus probability is a plausible explanations for these findings because they were experimentally controlled. These results support the emotional value hypothesis, according to which classical P300 processes reflect an affective evaluation of the stimulus, which in turn produces context updating.

Descriptors: Event-related potentials, P300, Emotion, Evaluative perception

Since the P300 component of the brain event-related potential (ERP) was first reported by Sutton, Braren, Zubin, and John (1965), considerable experimental effort has been directed towards uncovering its relationship to a variety of psychological processes. Task relevance and subjective probability have been shown to be the major antecedent variables that influence P300 amplitude (Johnson, 1989, 1993). In a task where a participant has to count the number of low-pitched tones presented in a random sequence with equiprobable high-pitched tones, P300 amplitude is significantly higher to the target stimulus (i.e., the task-relevant tone whose count has to be updated). Moreover, lower probability (less expected) target tones will elicit a larger amplitude P300 component than higher probability target tones (Duncan-Johnson & Donchin, 1977). These effects of subjective probability and task relevance have been observed in a number of different experimental tasks such as guessing, feedback, and signal detection paradigms (e.g., Fabiani, Gratton, Karis, & Donchin, 1987).

In addition to the conditions that elicit it, a further defining characteristic of an ERP component is its functional consequences. It has been shown that P300 processes are an index of stimulus evaluation that is separable from response selection and execution processes (McCarthy & Donchin, 1981). Furthermore, P300 amplitude is predictive of memory performance when experimental participants do not engage in cognitive elaborations of stimulus materials (Karis, Fabiani, & Donchin, 1984). On the basis of these results, Donchin and Coles (1988) proposed a context updating model that considers the P300 component to be a manifestation of neural activity occurring whenever the model of the environment in working memory must be revised.

A subfield of research on P300 amplitude has focused on examining the generality of the task relevance dimension. In the classical paradigm, the task relevance of a stimulus was defined by specific task instructions. Johnson and Donchin (1978) showed that P300 increases as a function of how much the subject relies on a given stimulus for correct task performance. A second class of experiments has used monetary rewards for manipulating the relevance of stimuli (Begleiter, Porjesz, Chou & Aunon, 1983; Johnston, 1979; Johnston & Holcomb, 1980). In these cases, the relevance that an experimental participant assigns to stimuli depends upon the learned value of money, and this value has generality beyond the experimental task. Under these conditions, the amplitude of P300 has been shown to systematically increase with monetary value. A third group of studies has employed stimuli that have intrinsic psychological relevance to experimental participants, even in the absence of task instructions. Johnston, Miller, and Burleson (1986) found a U-shaped function of P300 amplitude as a function of emotional value when participants were exposed to equally probable and task relevant slides that belonged to one of five categories ranging from the very unpleasant to the very pleasant (dermatological patients, neutral people, babies, and male and female nudes). Both pleasant and unpleasant slides evoked a larger P300 amplitude than did neutral slides. This U-shaped function has been replicated by other investigators (Mini, Palomba, Angrilli, &
Bravi, 1996; Naumann, Bartussek, Diedrich, Vogelbacher, & Mehrten, 1993). The manipulation of the internal state of the body (hormones, blood sugar level, salt/water ratio, etc.) constitutes another research strategy for changing the emotional value of stimuli. Johnston and Wang (1991) used the natural fluctuation of hormone levels in females and showed that identical pictures elicited different P300 amplitudes at different phases of their menstrual cycle in an adaptive manner. As an example, women in the high-progesterone phase elicited both larger P300 responses and higher pleasantness ratings to pictures of babies than did women in the low-progesterone phase.

Some further evidence of the need for a more generalized notion of task relevance has been provided by Farwell and Donchin (1991). They examined the P300s elicited by three different categories of stimuli (targets, probes, and irrelevants) in participants who had or had not been exposed to espionage scenarios on the day prior to testing. On the day of testing, participants were instructed to perform a classification task by pressing one switch when they saw a task relevant target stimulus (1/6 of the stimuli) and another switch for all non-target stimuli (irrelevants). However, probe stimuli, items taken from the espionage scenario on the prior day, were also presented (1/6 of the stimuli) in a random order during the classification task. As expected, target stimuli elicited larger P300s than did irrelevant stimuli, but probe and target stimuli were equivalent in the exposed group. That is, stimuli with which subjects had been involved on the day before were treated as task-relevant stimuli. They also elicited large P300s even when they had no relevance whatsoever for the experimental task.

Although the classical model of the P300 component accommodates these results by making them contingent upon the specific categorization processes used in each case, it does not provide a rationale for making a priori predictions on P300 amplitude effects evoked by stimulus dimensions considered relevant by participants but not by task instructions (Farwell & Donchin, 1991). An alternative theoretical perspective has considered emotional value as a more basic construct underlying the classical antecedents to P300 elicitation (Begleiter et al., 1983; Johnston et al., 1986). According to this more general model, a stimulus will be considered relevant if it is emotionally significant to the subject, whether that significance is established by task instructions or by his/her internal states or past experiences. The term emotional value refers to the intensity or arousal properties of affective states in current scaling models of emotion (Russell, 1980; Watson & Tellegen, 1985) that would represent the subjective importance of affective states of either positive (appetitive) or negative (aversive) valences. From an adaptive viewpoint, emotionally significant stimuli have utility for the accomplishment of survival and reproductive functions. Because the perception of changes in the uncertainty of occurrence of emotionally valuable events would also be of personal importance, the model considers subjective probability as an aspect of stimulus relevance. Such an evaluation mechanism would be useful for monitoring meaningful environmental stimuli and as a signal for the acquisition of useful knowledge in the strategic planning of future behavior (Johnston & Holcomb, 1980). This theoretical perspective has heuristic value for the study of the mnemonic functional consequences of P300 elicitation (Adley, 1967; Donchin, 1981; Gold, 1987; Kahneman & Miller, 1986; Roth, 1983; Simonov, 1983). According to the emotional value theory, however, the affective encoding of a stimulus is reflected in P300 processes and leads in turn to adaptive context updating.

One problem in contrasting the classical and affective models of P300 amplitude has been that of stimulus control. Typical experiments of the first kind have used simple visual or auditory stimuli such as high- and low-frequency tones. Affective P300 eliciting paradigms have used more ecologically realistic stimuli such as pictures of babies. Even if the likelihood of presentation of the individual pictures and the different emotional categories of stimuli were controlled in previous experiments (Johnston et al., 1986), their physical characteristics were not. However, the uncontrolled experimental probabilities and real world probabilities of stimulus features could leave results open to alternative interpretations. The purpose of the present study was to contrast the classical and affective models of the P300 component in the domain of facial perception by eliminating these potentially confounding probability effects. Faces were used as stimuli for three reasons: (a) new image processing methodologies allow for experimental control of their physical characteristics (Caldwell & Johnston, 1991), (b) the possibility of manipulating affective value by varying stimulus configurations such as proportions (eye-to-chin distance) and the fullness of lips that have been shown to affect perceived attractiveness (Johnston & Franklin, 1993; Perret, May, & Yoshikawa, 1994) and (c) the subjective relevance of facial stimuli in everyday life, which includes many different kinds of social transactions from mother-offspring communication to sexual selection and the monitoring of emotions in other individuals.

Previous studies using faces as stimuli have led to conflicting results. Although in some cases neutral faces evoked lower P300 amplitudes than did emotional faces (Carretie & Iglesias, 1995; Laurian, Bader, Lanares, & Oros, 1991), in other cases they evoked the highest (Vanderploeg, Brown, & Marsh, 1987). Gender effects were not systematically analyzed in either of these cases, and stimuli were characterized by being wholistic presentations of complex pictures or a reduced number of simple line drawings. The methodology used in the present study allows for testing of specific feature configuration such as short jaw proportions and full lips on the basis of previous attractiveness studies and normative data (Farkas, 1981; Johnston & Franklin, 1993). Both the facial stimuli and their physical feature components were equiprobable throughout the experiment. They also were equally task relevant because each participant was instructed to carefully examine the stimulus slides. Attractiveness ratings were postponed until the ERP collection sessions were complete. This procedure provided a measure of stimulus evaluation and eliminated possible response selection confounds. Under these conditions, the classical P300 model does not have theoretical elements for deriving how subjects are going to categorize relevance of stimulus dimensions beyond experimental instructions. It would therefore make no a priori predictions on gender differences or on differences among faces in P300 amplitudes. In contrast, the emotional value theory would state that relevance will be established by affective categorization of facial beauty or handsomeness. It also predicts, in both male and female participants, an association between P300 amplitude and subjective attractiveness, a dimension not included in the experimental instructions.

Method

Participants

Twenty-two male and 40 female undergraduate students participated in the experiment. Their ages ranged from 18 to 25 years, and they were given credit toward the requirements of an introductory psychology course. Female volunteers were required to (a) have a regular menstrual cycle, (b) not have used contraceptive pills within the preceding 3 months, (c) not be currently breast-
feeding a child, and (d) be more than 7 years after their menarche. All subjects were required to sign a consent form that described the contents of the slides and the procedure for attachment of scalp electrodes. The instructions also informed them that if they objected to the contents of the slides or had a homosexual preference they should withdraw from the experiment without any penalty.

**Stimuli**

FacePrints software (Caldwell & Johnston, 1991) was used to develop 32 female and 32 male gray-scale facial images. Two levels of proportions (short, average), hair (black, blond), eyes (narrow, wide), lips (narrow, full), and chin (narrow, wide) were used to generate a complete factorial combination of the six dichotomies (see Figure 1). The configuration of the proportions feature was based on prior studies establishing measurements for average and attractive female faces. The empirically derived short and average jaw proportions for the male faces provided experimental control for possible real-world probability effects that could influence the female ERP data. That is, if low real-world probability was the relevant variable, then larger P300s should be elicited by the less probable short-jaw proportions of both male and female faces. A Kodak carousel projector fitted with an electronic shutter back-projected the stimulus slides onto a translucent plexiglass screen. The images subtended approximately 10 degrees of visual angle.

**Apparatus**

Electroencephalogram signals were amplified using Grass Model 7P122 DC amplifiers with upper and lower half-amplitude frequency responses at 60 Hz and 0.04 Hz, respectively. The amplified signals were recorded by an IBM/XT computer fitted with a Tecmar Labmaster board for analog to digital conversion. ERPs were recorded and stored on the IBM hard disk between stimulus presentations. The computer, using the Tecmar Labmaster board, also controlled the slide projector (used to present the stimuli), the button panel (used by the subjects to initiate a trial and for responding during the fourth experimental session), and the tone generator (used to signal beginning and end of a trial).

**Procedure**

Each subject was fitted with three active nonpolarizing (Ag/AgCl) electrodes at the midline frontal (Fz), central (Cz), and parietal (Pz) locations designated by the International 10–20 system (Jasper, 1958). Subjects wore two linked reference electrodes placed on the earlobes and two electrooculograph (EOG) electrodes diagonally above and below the left orbit. The EOG electrodes were used to monitor vertical eye movements. Data from the EOG channel were used to reject any ERPs contaminated with eye movement potentials or blinks. After subjects were fitted with the electrodes, they sat in an electrically shielded, semidark room, with a projection screen directly in front of them and the button panel adjacent to their right hand. At this point, the resistance on each channel was checked and electrodes were adjusted until that resistance was less than 10 kΩ.

Following the test procedures participants were instructed that they would hear a 500-Hz tone for 100 ms at the beginning of each trial. This tone indicated when subjects could press a button to initiate the presentation of a stimulus slide. After a 1,000-ms pause, one of the picture slides was projected for a duration of 1 s. At that point, the projector advanced to the next slide and the tone indicated that the next trial could now be initiated by a button press. Subjects performed no experimental task other than viewing the stimulus material, although they were instructed to pay close attention to the slides because some questions would follow the third experimental session. ERPs were recorded for 1,070 ms, starting 50 ms prior to the onset of each visual stimulus, and were digitized at a rate of 100 samples/s.

In each experimental session, subjects were exposed to the complete set of 64 faces in a random order. There were a total of three sessions, with a 5-min break between them. Each of the faces was therefore presented three times within the ERP recording period, which provided sufficient data for signal averaging in the analysis of main and two feature interaction effects. Participants were not informed about the purpose of the study until debriefing. At the end of the third session, each subject was instructed to rate each of the 64 faces on a handsomeness/beauty Likert scale ranging from 1 (least beautiful or handsome) to 5 (most beautiful or handsome). No ERPs were recorded during this behavioral rating session. Female participants were asked to report the first day of their next menstrual cycle. This information was used to analyze the effect of menstrual status, which has been previously shown to affect P300 amplitudes to real-world stimuli (Johnston & Wang, 1991).

**Results**

**Behavioral Data**

Attractiveness ratings were analyzed by a mixed model design with gender as a between-subject variable and face sex, proportions, hair, eyes, lips, and chin as within-subject variables. Some general patterns of gender differences and strong facial feature effects on the beauty/handsomeness ratings were found across
participants. Table 1 lists all significant effects. Overall female subject ratings ($M = 3.01, SD = 0.98$) were higher than overall male subject ratings ($M = 2.57, SD = 1.17$), and female faces were rated higher across subjects ($M = 3.10, SD = 1.05$) than were male faces ($M = 2.61, SD = 1.04$), as revealed by main effects of gender and face sex, respectively. A significant Gender × Face Sex interaction indicated that the magnitude of the difference between male and female faces was larger for male participants than for female participants. Facial feature effects that appear stronger provide a basic context to understand the influence of finer stimulus configurations.

### Proportions
Short facial proportions were considered more attractive ($M = 2.94, SD = 1.09$) than average proportions ($M = 2.78, SD = 1.06$) in female faces. However, a small difference in the opposite direction was observed for male faces, as indicated by a Face Sex × Proportions interaction. The main effect of proportions was due to the weight of female face ratings in the overall averages.

### Hair
Blond-haired faces were rated as more attractive ($M = 3.22, SD = 1.02$) than black-haired faces ($M = 2.50, SD = 1.00$) for both genders, resulting in a main effect of hair. The difference was in the same direction but was larger for male subjects than for female subjects, as indicated by a Gender × Hair interaction.

### Eyes
A statistically significant Face Sex × Eyes interaction revealed that open eyes made female faces more beautiful ($M = 3.25, SD = 1.07$) than did narrow eyes ($M = 2.96, SD = 1.07$) but not more handsome in the case of the male stimuli. A Face Sex × Eyes × Hair interaction indicated that this advantage of open eyes in female faces was more pronounced on black-haired faces than on blond-haired faces, but the pattern of eye effects on male faces was not affected by hair type.

#### Lips
Full lips were generally rated as more beautiful or handsome ($M = 2.96, SD = 2.96$) than narrow lips ($M = 2.76, SD = 1.07$), as revealed by a main effect of lips. This preference for full lips was stronger for female than for male participants (Gender × Lips interaction).

The increased attractiveness of full lips over narrow lips across participants was strongest when full lips were in conjunction with open eyes rather than narrow eyes (Eyes × Lips interaction). This pattern of results was due to the influence of female face stimuli; it did not occur on the male face data (Face Sex × Eyes × Lips interaction). The advantage of full lips versus narrow lips was also larger in the context of average proportions than on faces with short proportions (Lips × Proportion interaction).

### Chin
This is the feature whose effect on the attractiveness ratings was most different for male and female faces. Although the mean preference for broad chins was significantly larger than for narrow chins, this was mainly a male face effect. A Face Sex × Chin interaction clarified the situation by showing that both simple chin effects were significant but acting in opposite directions. That is, subjects rated female faces with narrow chins as more beautiful,

### Table 1. F ratios for the Repeated Measures Analyses of Variance for Attractiveness Ratings and P300 Amplitudes

<table>
<thead>
<tr>
<th>Effect</th>
<th>Rating</th>
<th>P300 peak</th>
<th>Factor 2 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>16.47***</td>
<td>12.34***</td>
<td>24.78***</td>
</tr>
<tr>
<td>Face sex</td>
<td>33.78***</td>
<td>8.34**</td>
<td>6.51*</td>
</tr>
<tr>
<td>Proportions</td>
<td>21.12***</td>
<td>1.06</td>
<td>2.41</td>
</tr>
<tr>
<td>Hair</td>
<td>99.42***</td>
<td>4.89*</td>
<td>7.91**</td>
</tr>
<tr>
<td>Eyes</td>
<td>12.73***</td>
<td>0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>Lips</td>
<td>20.36***</td>
<td>4.50*</td>
<td>8.30**</td>
</tr>
<tr>
<td>Chin</td>
<td>6.60*</td>
<td>1.91</td>
<td>1.76</td>
</tr>
<tr>
<td>Gender × Face Sex</td>
<td>5.46*</td>
<td>2.64</td>
<td>5.00*</td>
</tr>
<tr>
<td>Gender × Hair</td>
<td>4.02*</td>
<td>1.73</td>
<td>4.71*</td>
</tr>
<tr>
<td>Gender × Lips</td>
<td>5.31*</td>
<td>2.30</td>
<td>0.27</td>
</tr>
<tr>
<td>Face Sex × Proportions</td>
<td>47.51***</td>
<td>3.47</td>
<td>2.22</td>
</tr>
<tr>
<td>Face Sex × Eyes</td>
<td>9.41**</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Face Sex × Chin</td>
<td>24.07***</td>
<td>0.00</td>
<td>2.36</td>
</tr>
<tr>
<td>Proportions × Lips</td>
<td>37.48***</td>
<td>0.30</td>
<td>4.85*</td>
</tr>
<tr>
<td>Proportions × Chin</td>
<td>14.18***</td>
<td>7.68**</td>
<td>7.32**</td>
</tr>
<tr>
<td>Eyes × Lips</td>
<td>9.25**</td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>Gender × Face Sex × Proportions</td>
<td>0.23</td>
<td>6.85*</td>
<td>5.30*</td>
</tr>
<tr>
<td>Gender × Proportions × Lips</td>
<td>1.96</td>
<td>1.56</td>
<td>6.71*</td>
</tr>
<tr>
<td>Gender × Proportions × Chin</td>
<td>4.41*</td>
<td>0.59</td>
<td>0.75</td>
</tr>
<tr>
<td>Gender × Eyes × Mouth</td>
<td>0.65</td>
<td>5.88*</td>
<td>1.72</td>
</tr>
<tr>
<td>Face Sex × Proportions × Chin</td>
<td>23.24***</td>
<td>0.04</td>
<td>0.87</td>
</tr>
<tr>
<td>Face Sex × Hair × Eyes</td>
<td>5.79*</td>
<td>1.21</td>
<td>0.04</td>
</tr>
<tr>
<td>Face Sex × Eyes × Lips</td>
<td>20.31***</td>
<td>0.04</td>
<td>0.88</td>
</tr>
</tbody>
</table>

**Note:** All tests without superscripts were conducted at the parietal site, $df = 1.60$. Effects not meeting a .05 significance criterion have been omitted.

*Main effect test, with maximal parietal effect and $df = 1.60$. *Electrode site interaction test, ($df = 2.120$) with Huynh and Feldt correction.

*p < .05, **p < .01, ***p < .001.
Fig. 2. Average waveforms for male and female subjects at parietal (Pz), central (Cz), and frontal (Fz) locations.

(M = 3.15, SD = 1.05) than those with broad chins (M = 3.06, SD = 1.06), whereas male faces with broad chins were considered more handsome (M = 2.76, SD = 1.07) than those with narrow chins (M = 2.47, SD = 0.99).

Significant Proportion × Chin and Gender × Proportion × Chin interactions qualified the chin main effect described above. They revealed that the overall preference of broad over narrow chins was significant in the rating of faces with short proportions but not the rating of faces with average proportion. This result was observed only on female participants and not on male participants, where the two-way Proportion × Chin interaction was not significant.

An analysis of the significant Face Sex × Proportions × Chin interaction indicated that simple Proportion × Chin interaction effects occurred only in the ratings of male faces and not the ratings of female faces. More specifically, the increased attractiveness of broad chins in males was much larger on short proportion ratings of female faces. More specifically, the increased attractive-ness of broad chins in males was much larger on short proportion than on average proportion faces.

Electrophysiological Data

Average waveforms for male and female participants at frontal, central, and parietal locations are shown in Figure 2. For each subject, the data set contained 192 average waveforms (2 Face Sex × 2 Hair × 2 Proportions × 2 Eyes × 2 Lips × 2 Chins × 3 Electrode Sites). The data from 22 male subjects and 40 female subjects generated a total of 11,904 ERPs. After baselines were subtracted from these individual ERPs, a principal component analysis was performed on their covariance matrix using the Varimax method. After rotation, the first three factors accounted for 26%, 22%, and 16% of the variance, respectively. The temporal and spatial distribution of these factors was used to associate them with ERP components (Figure 3). Factor 1 (F1) was positive, started late, and loaded maximally at 900 ms after stimulus onset; it had a Fz > Cz > Pz scalp distribution and was identified as the frontal slow wave (Naumann, Maier, Diedrich, Becker, & Bartussek, 1997; Ruchkin, Johnson, Mahaffey, & Sutton, 1988). Factor 2 (F2) was positive, with a maximum loading at around 550 ms after stimulus onset. Its characteristic Pz > Cz > Fz scalp distribution, along with previous studies reporting oddball effects at similar latencies using complex visual stimuli (Naumann et al., 1993), led to its tentative identification as the P300 component. Factor 3 loaded maximally in the vicinity of 150 ms after stimulus onset. Peak measures that correspond to the tentative P300 component were taken as the average amplitude (±50 ms) around the most positive peak in the 450–650-ms time window. Slow wave area measures were taken as the average amplitude from 650 ms to the end of the recording epoch.

Facial Feature Effects

ERP data were analyzed by a mixed model design with gender as a between-subject variable and face sex, proportions, hair, eyes, lips, chin and electrode site as within-subject variables. Overall mixed model analyses of variance on both F2 scores and corresponding peak scores revealed the typical Pz > Cz > Fz pattern of P300 effects as a function of facial feature configurations. Most statistically significant experimental effects of this component appeared at the parietal site and disappeared gradually from central to frontal sites. Therefore, the description of this component will focus on the maximally sensitive parietal site analysis. All F2 effects discussed here were statistically significant at the parietal site, as is shown in Table 1 along with corresponding effects on peak amplitudes and behavioral ratings.

Male participants had a much larger overall F2 score response to the facial stimuli (M = 0.75, SD = 1.18) than did female participants (M = 0.06, SD = 0.98), as revealed by an effect of gender (Figure 2). Female faces evoked larger F2 scores across participants (M = 0.36, SD = 1.10) than did male faces (M = 0.25, SD = 1.11), as indicated by an effect of face sex. The difference, although in the same direction, was larger for female than for male subjects, where the simple effects were not statistically significant (Gender × Face Sex interaction). Slow wave magnitude was also larger for male than for female participants. The difference was only statistically significant on area measures, F(1,60) = 8.07, p < .01, and not on F1 scores. No feature effects are described on the F1 component because there were no a priori predictions.

Proportions. A very different pattern of face sex and proportions effects on F2 for male and female subjects resulted in a three-way interaction of these two variables with gender. For male participants, male faces with average proportions elicited larger F2 scores (M = 0.85, SD = 1.18) than did male faces with short proportions (M = 0.63, SD = 1.14). In contrast, female faces with short proportions evoked larger F2 scores (M = 0.80, SD = 1.19) than female faces with average proportions (M = 0.72, SD = 1.18). This cross-over of effects was absent in female participants, whose F2 response was larger to female than to male faces (simple effect...
of face sex) and did not show a significant two-way Face Sex × Proportions interaction.

Hair. Blond haired faces elicited larger F2 scores ($M = 0.34$, $SD = 1.13$) than did black-haired faces ($M = 0.26$, $SD = 1.08$), as indicated by a main effect of hair. This average effect was mainly due to the simple hair effects in male participants because the simple hair effects in female subjects were not significant (Hair × Gender interaction).

Lips. A significant effect of lips was observed; full lips elicited larger average F2 scores ($M = 0.34$, $SD = 1.10$) than did narrow lips ($M = 0.26$, $SD = 1.10$). However, this average pattern of effects was mainly characteristic of female participants, as revealed by a Gender × Proportions × Lips interaction. For male participants, the increase in F2 scores for full lips occurred only in faces with short proportions. The effect of lips on average-proportion faces had the opposite direction in male participants: narrow lips evoked larger F2 scores than full lips. This cross-over interaction was absent in female subjects, who showed the former main effect advantage of full lips for both proportion levels.

Chin. Broad chins elicited larger F2 scores than narrow chins but only on short-proportion and not on average-proportion faces, as revealed by a significant Proportions × Chin interaction.

Correlation Between Attractiveness Ratings and Late Positive Components

Average ratings and ERPs were computed over all male and all female participants for each male and female face, generating a data set of 64 average waveforms and ratings for each gender.

Male participants. A strong correlation between male F2 scores and beauty ratings of female faces was found at the parietal site ($r = .40$, $p < .009$) (see Figure 4). A parallel effect was found between peak amplitudes and ratings ($r = .44$, $p < .01$). No significant association effects between F2 scores and handsomeness ratings were found for the male faces, although there was a trend in the same direction ($r = .26$).

Female participants. The association between attractiveness ratings and F2 scores was weaker for female subjects. Although correlation trends in the same direction were observed for female faces ($r = .29$) and male faces ($r = .17$), neither of them was statistically significant. However, regression lines between F2 scores and attractiveness ratings for male and female participants were close to parallel in each face sex, and gender differences in their corresponding slopes were not statistically significant for either male or female faces. Similar results were obtained on peak measures. An analysis of the effects of menstrual phase was performed next because the hormonal changes involved have been shown to affect P300 processes (Johnston & Wang, 1991).

Although the duration of the menstrual cycle varies among women, this variance is almost exclusively confined to the pre-ovulatory phase. Thus, ovulation is almost exactly 14 days prior to the onset of the next menses, irrespective of the cycle length (Fluhmann, 1957; Lein, 1979; Matsumoto, Nogami, & Ohkuri, 1962). In a 28-day cycle, ovulation occurs on about the 14th day, but in a 34-day cycle ovulation occurs on about the 20th day (Katchadourian, 1980). Therefore, participants were asked to report when the first day of the next menses occurred. The estimated menstrual cycle day for each woman at the time of the experiment was then determined by counting backwards from the day that was reported.

The estimated menstrual cycle day was used to classify each woman into one of three menstrual phases, defined on the basis of usual blood concentration levels of steroid hormones (Judd and Yen, 1973; Thorneycroft, Mishell, Stone, Khanna, & Nakamura, 1971). Ten women were assigned to each of these three phases: preovulatory (before day 8, when levels of estrogen, progesterone, and androgen are at their minimum), ovulatory (days 11–18, when levels of androgen are at their maximum), and postovulatory (days 19–24 when progesterone levels are at their peak and estrogen levels are high). The remaining women were not included in the menstrual phase analyses either because of missing menses reports or because they could not be unambiguously classified into one of the above groups.

In the preovulatory group, a frontal correlation between handsomeness ratings and P300 measures was observed for both F2 scores ($r = .41$, $p < .02$) and peak measures ($r = .39$, $p < .03$). No
statistically significant association between these two variables occurred for the ovulatory group.

For the postovulatory group, a significant association between P300 measures and handsomeness ratings was observed at the parietal site for F2 scores ($r = .44, p < .02$), but the corresponding peak effect was smaller ($r = .25, p < .2$). A strong and widespread correlation across electrode sites was also observed between F2 and beauty ratings ($r = .55, p < .002$ at Fz; $r = .64, p < .001$ at Cz; and $r = .61, p < .001$ at Pz). Peak results reflected an even stronger association ($r = .78, p < .001$ at Fz; $r = .72, p < .001$ at Cz; and $r = .64, p < .001$ at Pz). In this group, slow wave area measures also showed a correlation with beauty ratings at the central site ($r = .49, p < .005$).

ERP differences between women in the two most extreme menstrual phases and hormonal profiles were associated with rating differences. Overall average ratings were less consistent between postovulatory and preovulatory women ($r = .57$) than between postovulatory and ovulatory women ($r = .78$). The difference between these two correlation values was statistically significant ($Z = 2.19, p < .05$) and mainly due to changes in the ratings of female faces.

**Discussion**

The facial feature manipulations were effective in producing differences in subjective attractiveness of the stimuli. They replicated...
Gender differences in late positive components

Results were more alike between males and females than were ERP responses from the behavioral ratings of male and female subjects were revealed that the concept of beauty and handsomeness were quite observed for either male or female faces.

In male participants, short proportions elicited larger P300s for both male and female participants, although in the women it was qualified by their menstrual phase. Although there were no specific predictions on the effect of menstrual phase, these data fit well with the affective viewpoint. Mood changes over the menstrual cycle are well documented, and previous research has shown that P300s to emotional stimuli vary with menstrual phase (Johnston & Wang, 1991). The linear rather than U-shaped function effects of attractiveness that were found here may be due to the lower arousal properties of unattractive faces in comparison with aversive stimuli, such as skin diseases, used in previous studies. These results lend support to the hypothesis that the P300 component reflects an affective encoding of stimulus materials. These results are difficult to explain on the basis of the classical P300 model because stimuli were equally probable and task relevant and the theory does not have a priori predictive scope when relevant stimulus dimensions, such as attractiveness, are not explicit in task instructions.

But, is the F2 a measure of the classical P300 component? Several sources of evidence suggest that it is. First, F2 effects had the classical P300 scalp distribution. Even when some association existed between F2 and the ratings at the frontal site in female subjects, the magnitude of stimulus effects was larger at the parietal site. Second, it was a measure of stimulus evaluation not dependent on response selection processes, because the task of the subject was simply to view each face. Third, its latency replicates previous findings with complex visual stimuli where oddball effects were reported (Naumann et al., 1993). Fourth, the predictability of F2 scores from attractiveness ratings brings these results in line with other studies from affective paradigms using simpler stimuli where the P300 component has been observed in the classical time window from 300 to 400 ms.

Both the gender differences on the ERPs associated with stimulus characteristics and those associated with physical facial feature effects are difficult to explain on the basis of the classical antecedent variables, which were experimentally equated across conditions. A possible criticism might arise by arguing that even though all features were equally probable within the experimental situation, they might have different real world probabilities. An analysis of face sex effects shows that this explanation is not plausible. Because the probability of finding male or female faces in the real world is almost equal, they would be expected to elicit similar P300 amplitudes. Female faces however evoked larger responses than did male faces and especially so in female subjects. Second, average proportions have higher frequencies than short-jaw proportions for both male and female faces. If probability effects were to be assessed with reference to the real world, then short proportions should elicit larger P300s than average proportions irrespective of the sex of the face. This, however, was not the case. In male participants, short proportions elicited larger P300s than did average proportions in female faces but smaller P300s in male faces. In female participants, no effect of proportion was observed for either male or female faces.

The occurrence of numerous face sex interaction effects revealed that the concept of beauty and handsomeness were quite different. However, the concepts of beauty and handsomeness measured from the behavioral ratings of male and female subjects were very similar; they differed only in degree. Generally, behavioral results were more alike between males and females than were ERP results. Although most of the P300 effects had behavioral counterparts, the dissociations in gender differences observed between these two psychological measures revealed that they constitute different tools in the study of facial processing. If the P300 response is viewed as an index of the participant emotional response, then the following interpretations can be given. First, the affective value evoked by the stimulus faces was larger for male than for female participants. Second, both males and females responded on average more to female than to male stimulus faces, but the difference in affective value was most pronounced in female participants. With respect to the proportions of a female face, the previous cross-cultural finding of a rating preference for short over average jaw length was replicated. This behavioral result was observed in both male and female participants, but according to the ERP findings only male participants had a larger affective response to the shorter jaw configuration of female faces. In this case, it appears that the P300 response is tapping into some evaluative perception processes not reflected by the ratings. The behavioral scale used in the current study provided a measure of attractiveness, only one of several possible emotional dimensions. Use of alternative behavioral scales, such as eroticism or arousal, may improve prediction of P300 amplitude.

The direction of gender effects described above is in sharp contrast to that found in a previous study. Polich and Martin (1992) found a gender main effect and an interaction between gender and probability in a classical auditory oddball task. P300 amplitude observed in female subjects was larger than amplitudes in male subjects. The difference increased with decreasing probability of the target stimulus. Thus, the occurrence of this result may depend on stimulus or task characteristics. An analogous situation has been found in menstrual studies. Whereas no effect of menstrual phase on P300 amplitude was obtained with affectively neutral stimuli (Fleck & Polich, 1988), phase differences appeared when emotional stimuli were used (Johnston & Wang, 1991). According to the emotional value theory, specific differences in emotional structures and reactivities between males and females account for these gender effects. Females may experience a higher degree of surprise than males to unexpected events such as the target stimulus in the previous oddball task. In the present case, the larger P300 amplitudes observed in male subjects may reflect a higher sensitiveness to physical attractiveness cues (Berscheid & Walster, 1974; Ford & Beach, 1951). The larger P300 to female faces than to male faces observed in female subjects was unexpected but was consistent with attractiveness ratings. Although the saliency of female physical features to other female subjects has been pointed out as one of the basic currencies of intrasexual competition (Symons, 1979), this effect may depend on the particular set of faces used as stimuli. The female features in this study were derived from prior studies of female facial beauty. No such studies were available to guide the selection of male features. Further studies should evaluate this interpretation using different sets of pictures and investigate specific hypotheses derived from existing knowledge of male–female differences in affective responses (Durden-Smith & deSimone, 1983). Here again, the classical P300 model lacks theoretical elements for making a priori predictions of gender differences in the assignment of stimulus relevance to dimensions that are not explicit by task instructions.

Present data support the proposed construct of emotional value as a representation of stimulus relevance and as more parsimonious than the two traditional concepts of task relevance and subjective probability as antecedent variables in accounting for variance in P300. It is also more consistent with the notion of surprise as a
behavioral descriptor for P300 processes (Donchin, 1981). The degree of surprise elicited by an unexpected low-probability event appears to depend on its personal relevance to the individual. Consider somebody buying a lottery ticket and then discovering that he or she has won the prize. He or she will probably find this to be a more surprising event than hearing that an unknown person was the winner. The probability of any specific winner is the same, but the personal value for the particular buyer of the ticket is not. As a consequence, the degree of surprise elicited by events of low subjective probability would depend on its personal value. In these respects, the emotional value theory provides a more general formulation than the classical model of P300 amplitude, not only because of its wider predictive scope of its interpretation of stimulus relevance but also because its integrative consideration of the classical antecedents of P300 amplitude leads to novel hypotheses. According to the affective model, the impact on P300 amplitude of changes in the subjective probability of a stimulus would depend and not simply add to the impact of stimulus relevance (Johnson, 1989, 1993). The more general interactive model that results as a consequence encompasses the additive model as a special case under specific task conditions. Further studies varying the probability of emotional events may discriminate between these alternative positions.

The acquisition of new knowledge and the modification of internal representations (context updating) have been proposed as a functional consequence of P300 elicitation (Donchin, 1981). The emotional value hypothesis also fits well with this interpretation because it views P300 processes as an index of the significance of events that would drive an organism to adaptively update and store contextual information. P300 amplitude to predictive stimuli increased as a function of the monetary reward linked to the predicted stimulus in an associative learning paradigm (Johnston & Holcomb, 1980). Degree of surprise is also a good predictor of amount of conditioning in the associative learning literature (Lieberman, 1990). It signals the amount of change in the predictability of rewards and punishments that an organism experiences when exposed to unexpected stimuli. However, a general relationship between P300 amplitude and memory performance has not been found. Only under certain experimental conditions such as incidental learning paradigms or the use of surprising, physically distinctive words in tasks where the participant does not use elaborate strategies has P300 amplitude been accompanied by memory improvement (Fabiani, Karis, & Donchin, 1990). The specific nature of the way these representations are established is an open question. The emotional value theory provides a new insight into this problem by suggesting that the class of memory processes involved is related to the formation of affective memories (Ledaux, 1994). Prospective work on the association between memory performance and P300 amplitude to emotional stimuli may shed light on this question and on the specific nature and strategic role of the processing functions indexed by the P300 component.

REFERENCES


(Received September 4, 1997; Accepted August 10, 1998)