Metamemory, distinctiveness, and event-related potentials in recognition memory for faces

W. SOMMER, A. HEINZ, H. LEUTHOLD, J. MATT, and S. R. SCHWEINBERGER
University of Konstanz, Konstanz, Germany

A neglected topic in metamemory research is the ability of subjects to predict their own recognition performance for faces. We investigated whether subjects can make such judgments of learning (JOLs) for unfamiliar faces and whether JOLs relate to facial distinctiveness, a powerful determinant of face recognition. One group of subjects made JOLs, and a second group rated the same faces for distinctiveness; subsequently, both groups tried to recognize these faces among new faces. There was significant prospective metamemory for faces that appeared to be based on facial distinctiveness. Both prospective metamemory and distinctiveness ratings related to long-lasting effects in event-related brain potentials (ERPs), closely resembling an ERP component that predicted face recognition. Therefore, the brain processes underlying JOLs, distinctiveness, and recognition memory for faces appear to be intimately related.

Self-monitoring is considered to be essential for the regulation of cognition and behavior. Thus, knowledge about memory—that is, metamemory—may be important for the allocation of study time, for the termination of retrieval attempts, and in determining confidence in recall or recognition. Metamemory has been extensively investigated in verbal learning and memory (for a recent review, see Nelson & Narens, 1990) but has been largely neglected in a field where it may be of considerable practical relevance, the recognition of faces. Since face recognition does not seem to be mediated by verbal coding (e.g., Malpass, Lavigne, & Weldon, 1973), rules for facial metamemory cannot be deduced from findings on verbal metamemory. Considering the juridical weight given to eyewitness testimonies (see J. W. Shepherd, Ellis, & Davies, 1982), it is important to study the validity of the prospective and retrospective confidence of subjects to recognize a face in question.

Brigham (1988) and Deffenbacher (1980) have investigated retrospective confidence judgments of face recognition. Recently, Vokey and Read (1992) reported that memorability ratings for faces derived from one group of subjects are predictive for face recognition performance in a second subject group. However, there appear to be no studies of the relation between recognition predictions and recognition performance in the same subjects. In the context of verbal learning, such estimates have been termed judgments of learning (JOLs) (e.g., Devolder, Brigham, & Pressley, 1990; Leonosio & Nelson, 1990; Lovelace, 1984). Therefore, the first aim of the present study was to investigate whether subjects can make valid JOLs for faces.

In the verbal domain, several potential bases for making JOLs have been considered (Nelson, Gerler, & Narens, 1984). Trace access theories (e.g., Arbuckle & Cuddy, 1969) suggest that subjects may have privileged access to memory traces. Alternatively, JOLs may be based on the previous retrieval experience with a given item (King, Zechmeister, & Shaughnessy, 1980; Vesonder & Voss, 1985). But even in the absence of prior retrieval experience, JOLs may be quite accurate. Therefore, JOLs may be based also on the general characteristics of the material to be memorized, such as word frequency or imagery (Groninger, 1979), ease of learning (Underwood, 1966; Vesonder & Voss, 1985), or ease of processing estimates (Begg, Duft, Lalonde, Melinek, & Sanvito, 1989).

Face recognition is sensitive to a number of social and stimulus-related factors, such as gender, race, familiarity, and facial distinctiveness as opposed to typicality (for reviews, see Bruce, 1990, and J. W. Shepherd, 1981). For rating purposes, distinctiveness has been operationalized as the ease of spotting a face in a crowd (Valentine & Bruce, 1986a, 1986b) or the similarity to a typical face (Light, Kayra-Stuart, & Holland, 1979). A number of studies have shown that distinctive faces are recognized better than are typical faces (Bartlett, Hurry, & Thorley, 1984; Cohen & Carr, 1975; Ellis, J. W. Shepherd, Gibling, & J. Shepherd, 1988; Going & Read, 1974; Winograd, 1981).

We hypothesize that facial distinctiveness may serve as a basis for recognition predictions. For one, the oper-

Footnotes:

Portions of this paper were presented at the symposium for New Developments in Event-Related Potentials, Hannover, May 1991, and at the First European Congress of Psychophysiology, Tilburg, June 1991. This research was supported by the Deutsche Forschungsgemeinschaft (Grant So 177/2-2) and was conducted by A. Hein in partial fulfillment of the requirements for the Diploma in Psychology, University of Konstanz. We thank Shlomo Bentin, Arthur Glenberg, and Tom Nelson for their comments on an earlier draft of this paper, and Ursula Lommen and Petra Mikulics for their help in data collection. Correspondence should be addressed to W. Sommer, Fachgruppe Psychologie, Universität Konstanz, Postfach 5560 (D47), D-78434 Konstanz, Germany.
ositionalization of distinctiveness used in some studies—
 ease of spotting a face in a crowd—resembles a meta-
memory judgment. In addition, facial distinctiveness is
related to retrospective metamemory. Brigham (1987,
cited in Brigham, 1988) found the correlation between
recognition accuracy and recognition confidence to be
higher for distinctive faces than for typical faces. There-
fore, it was the second aim of the present study to in-
vestigate how JOLs for faces relate to facial distinctive-
ness and to identify the locus of these effects within the
cognitive system.

One approach to elucidate the locus of experimental
effects within the flow of information processing is by
recording event-related brain potentials (ERPs) that re-
late to cognitive processes. If a cognitive process (ma-
nipulated by an experimental factor) is reflected in an
ERP component, its temporal characteristics (latency,
duration) can be used as a chronometric measure of the
related cognitive process. The amplitude distributions
of ERPs over the scalp can be used to differentiate between
brain systems underlying cognitive processes (see John-
son, 1993). The scalp topography of an ERP component
depends on the location and orientation of its neural
generators. If ERPs elicited in different conditions also
differ in topography, they must by necessity originate from
different brain systems. On the assumption of isomor-
phism between brain and cognition, cognitive processes
may also be dissociated on the basis of their ERP topogra-
phies, because it appears unlikely that one and the same
process is performed in different neural substrates.

ERPs at initial presentation have been found to be pre-
dictive of subsequent memory performance (see Kutas,
1988; Paller, 1993). Most studies have demonstrated that
ERPs to subsequently recalled or recognized words were
electrically more positive than were ERPs to for-
gotten words. These ERP Differences predictive of
memory performance, operationally termed Dm by
Paller, Kutas, and Mayes (1987), are often widely dis-
tributed over the scalp and extend over several hundred
milliseconds. Recently, we have reported a Dm for unfa-
miliar faces (Sommer, Schweinberger, & Matt, 1991).
The face Dm began 200 msec following stimulus onset
and was bipolar with more positive frontal and more
negative posterior ERPs to subsequently recognized
faces, relative to ERPs to unrecognized faces.1

It is conceivable that perceptual or attentional vari-
ations during early stimulus processing stages may affect
both study-phase ERPs and subsequent memory perfor-
mance. Sommer et al. (1991), therefore, required a per-
ceptual discrimination during the study phase. Because
perceptual sensitivity was the same for subsequently rec-
nognized and unrecognized faces, recognition perfor-
ance and thus Dm appeared to be independent of vari-
ations in early processing stages and possibly specific
for memory processes. Sommer et al. (1991) suggested
that the face Dm may reflect encoding processes into
memory. In the present study, we attempted to replicate
the independence of Dm from perceptual and attentional
factors, and we investigated whether distinctiveness and/
or JOLs relate to similar brain systems as those produc-
ing the Dm by comparing the time course and scalp
topographies of the respective ERP effects.

METHOD

Subjects
Twenty-six subjects participated in this study; the data of 2 sub-
jects were lost due to technical problems. The remaining 24
subjects were assigned to two groups of 12 subjects (6 men and 6
women) each. Mean age was 26.8 years for Group 1 (range 23 to
36 years) and 25.7 years (range 22 to 36 years) for Group 2. All
subjects were right-handed, with average handedness scores (Old-
field, 1971) of 9.90 in Group 1 (range 5 to 1.00) and 8.78 in
Group 2 (range 1.6 to 1.00). All subjects had normal or corrected-
to-normal binocular visual acuity (equal to or better than 20/25
Snellen, as tested with a Bausch and Lomb Vision Tester). The subjects were paid DM 10 per hour plus a bonus of up to DM 10,
depending on recognition performance.

Stimuli
Black-and-white photographs of male college students with
neutral expressions and without particular features, such as spec-
tacles or beards, were video-digitized from college yearbooks. All
faces were turned slightly to the viewer’s right. In order to occlude
background and hairstyle, the portraits were fit into a vertical el-
lipse. These faces were presented at the center of a computer
screen. Viewing distance, controlled by a fixed chinrest, was 1 m;
height and width of the face stimuli was 9.0 x 8.0 cm, corre-
sponding to visual angles of 5.2” x 4.6”. The average luminance
of the faces was 14.0 cd/m². For the assessment of perceptual per-
formance during memory encoding, a number of the presented
faces were manipulated in “contrast” by brightening the 9 darkest
gray shades (of 16 in all) for one level. This manipulation in-
creased luminance by about 0.7 cd/m² and created the subjective
impression of a slight contrast reduction.

Procedure
The subjects were seated in an electrically shielded, sound-
attenuated chamber, where a masking noise of 38 dB(A) was pro-
vided by ventilation. Each trial, whether during study or recogni-
tion, began with a 60-msec warning tone [84 dB(A), 1500 Hz]. At
260 msec, a gray ellipse (luminance 16.3 cd/m²) of the same
size and outline as the face stimuli appeared for 1,900 msec. The pre-
sentation of the ellipse was interrupted 840 msec after its onset by
one of the faces for 160 msec. The first rating was requested
500 msec after the disappearance of the ellipse.

All ratings were made on 4-point scales. For each type of rating,
the same horizontal bar (20.0 x 3.3 cm, 5.3 cd/m²) was presented
on the screen, subdivided by vertical lines into four equal rectan-
gles. The required rating was designated in writing above the scale.
For contrast ratings, the German words for flat and clear were
written above the leftmost and rightmost rectangles of the bar,
which represented the extremes of the rating continuum; for recog-
nition ratings, the German words for old and new represented these
extremes. For JOLs and distinctiveness ratings, the words yes and
no appeared above the left- and rightmost rectangles, and the words
recognize or distinctive were shown centered above the scale.
The rating was performed by selecting one of the four rec-
tangles with the mouse cursor of the Amiga. Upon selection, the
chosen rectangle turned red for 0.5 sec and the cursor disappeared
until the next rating. If the rating required more than 2 sec, a re-
minding tone [56 dB(A), 1000 Hz] was given. Successive ratings,
as required during study, were separated by an interval of
100 msec. Three seconds after the last rating the next trial began.
Seventeen pairs of stimulus block were presented. The first block
in each pair served for study and was followed by a recog-
nition block. The first pair of blocks served practice purposes and was not analyzed. In each block, 11 faces were shown. In the study blocks, five normal photos were randomly mixed with five contrast-reduced portraits. The 11th photo was a filler item. During study, after each photo, two ratings had to be given in succession. The first rating always concerned contrast, whereas the second one was a JOL in Group 1 and a distinctiveness rating in Group 2. In the recognition blocks, the first photo was a filler item. The other 10 portraits showed the five unmanipulated faces from the previous study block, randomly mixed with five new and unmanipulated portraits. Here, each face was rated from old to new.

The task for the filler items was the same as for the others, but all respective responses were discarded. Filler items were used to prevent recognition on the basis of sensory memory. The interval between study presentation and recognition test of nonfiller items ranged from 15 to 90 sec. The same sequence of stimuli was used for all subjects.

Questionnaire

Immediately after the experiment, subjects of Group 1 answered a questionnaire as to how frequently they had used particular heuristics in predicting face recognition (4-point scales from never to very frequently) and which had been useful and useless (yes/no) in actually recognizing the face. The following heuristics were suggested: face recognizability at presentation, difficulty of recall, difficulty to memorize, distinctive facial features, sympathy or antipathy, similarity to known persons, or any others. The same categories, except for the difficulty to memorize, were used in a questionnaire for Group 2, where the usefulness of these heuristics for recognition should be indicated.

Behavioral Data Analysis

Contrast discrimination and recognition performance was assessed by determining the areas below the ROC curves [P(A); Green & Swets, 1966]. In measuring contrast sensitivity, we considered the unmanipulated faces as signals and the contrast-manipulated faces as noise. In assessing recognition performance, old and new faces were taken as signals and noise, respectively. Statistical tests were based on arcsin-transformed data; means are given untransformed.

Associations between ratings were assessed with nonparametric gamma correlations (Goodman & Kruskal, 1954), as recommended for metamemory studies (Nelson, 1984). Matched partial correlations coefficients, excluding ties, were calculated with the method described by Quade (1974).

Electrophysiological Recordings

Initially, the EEG from 15 electrodes (Fz, Cz, Pz, F3, F4, C3, C4, P3, P4, O1, O2, T3, T4, T5, T6; see Figure 1) and the EOG from below the outer canthus of the right eye were commonly referenced to Fpz (bandpass 0.03 to 40 Hz). Electrode impedances were kept below 10 kΩ. The EEG and EOG were recorded with a sampling rate of 100 Hz for an epoch of 1.2 sec, starting 200 msec prior to stimulus onset. Artifact-free ERPs (EOG activity < 62.5 μV) were sorted and averaged according to the following conditions. First, ERPs to faces of normal contrast, presented during the study phase, were separated according to whether or not the face had been subsequently recognized. Second, the study phase ERPs to all faces were sorted according to the ratings during study—that is, positive versus negative JOLs (Group 1) and high versus low distinctiveness (Group 2)—and according to whether or not the photo had been contrast-reduced. Third, the

Figure 1. Topographical display of ERPs to “clear” faces in the study phase, averaged and superimposed according to subsequent recognition. Left and right panels present ERPs of Group 1 and 2, respectively, which either made judgments of knowing or rated facial distinctiveness.
ERPs from the recognition phase were averaged separately for old and new faces and according to whether they were rated as old or new.

For averaging purposes, cutpoints were defined for each kind of rating and each subject, dividing the frequency distributions of artifact-free ERPs over the 4-point scales as far as possible into dichotomous bins. After averaging, the waveforms were low-pass filtered at 8.8 Hz (~3 dB; Ruchkin & Glaser, 1978) and recalculated to an average reference derivation (Lehmann, 1987).

RESULTS

Recognition performance \([P(A)]\) was better than chance in both the JOL group \([M = .715, SEM = 0.015, t(11) = 13.77, p < .001]\) and the distinctiveness group \([M = .717, SEM = 0.024, t(11) = 8.86, p < .001]\); groups did not differ \([t(22) = 0.13]\).

Similarly, contrast sensitivity was above chance in both the JOL group \([M = .659, SEM = 0.023, t(11) = 6.98; p < .001]\) and the distinctiveness group \([M = .630, SEM = 0.023, t(11) = 5.89, p < .001]\); again, groups did not differ \([t(22) = 0.94]\). Correlations with contrast ratings could not be calculated for one subject of the distinctiveness group because only one rating category was used. Contrast and recognition ratings were slightly correlated in the JOL group \([G = -.16, SEM = 0.06, t(11) = -2.76, p < .05]\) but not in the distinctiveness group \([G = .11, SEM = 0.10, t(10) = 1.04]\).

Of central interest here is the relationship of the JOLs and distinctiveness ratings during study with recognition ratings during the recognition phase. Gamma correlations between these ratings were calculated for each subject separately. On average, JOLs and recognition ratings were positively correlated \([G = .44, SEM = 0.05, t(11) = 9.73, p < .001]\), as were distinctiveness and recognition \([G = .36, SEM = 0.05, t(11) = 7.47, p < .001]\). These correlations did not differ significantly \([t(22) = 1.27]\).

Whereas contrast ratings were, at most, weakly related with recognition, they were negatively correlated with both JOLs \([G = -.50, SEM = .09, t(11) = -5.79, p < .001]\) and distinctiveness ratings \([G = -.36, SEM = .09, t(10) = -4.02, p < .01]\). These correlations were not significantly different \([t(21) = -1.14]\).

If recognition predictions for a particular face are based on perceived distinctiveness, we should expect a correlation between the two types of ratings over faces. This hypothesis was tested by calculating median recognition predictions of Group 1 and median distinctiveness ratings of Group 2 for each of the 80 unmanipulated faces. Gamma between JOLs and distinctiveness was \(G = .75\) (asymptotic standard error \(ASE = .086, p < .001\)). Similarly, we calculated median recognition ratings for each face on the basis of ratings from both groups. Recognition correlated positively with distinctiveness \((G = .56, ASE = .093, p < .001)\) and with JOLs \((G = .53, ASE = .123, p < .001)\). Median contrast ratings were not significantly correlated with recognition \((G = -.13, ASE = .144)\), JOLs \((G = -.32, ASE = .158)\), or distinctiveness \((G = -.20, ASE = .145)\).

The correlation between distinctiveness and JOLs was not greatly changed when recognition ratings were controlled for by partial correlations \([G (JOL, distinctiveness | recognition) = .66, ASE = .106, p < .001]\) and also when contrast ratings were partialled out \([G (JOL, distinctiveness | contrast) = .75, ASE = .102, p < .001]\). Similarly, the correlation of recognition with distinctiveness was unaffected by controlling JOLs \([G (distinctiveness, recognition | JOL) = .51, ASE = .102, p < .001]\) or contrast \([G (distinctiveness, recognition | contrast) = .61, ASE = .093, p < .001]\). Finally, the correlation between JOLs and recognition was unaffected by partialling out contrast \([G (JOL, recognition | contrast) = .49, ASE = .145, p < .145]\) but was reduced to insufficiency when distinctiveness was controlled for \([G (JOL, recognition | distinctiveness) = .26, ASE = .168]\).

Questionnaire Reports

Table 1 presents the results of the postexperimental questionnaire for the JOL group. The heuristic most widely used in making recognition predictions was attending to distinctive facial features, closely followed by estimating the difficulty of memorizing. The heuristics used least were sympathy or antipathy toward the face and its similarity with known persons. The results about the usefulness of these heuristics in face recognition were even clearer. Distinctiveness of facial features was far more often considered useful than any other heuristic but never useless.

The questionnaire reports of Group 2, who rated facial distinctiveness during study, confirmed the outstanding importance of this dimension in recognition. Distinctiveness was considered useful 12 times, which is twice as often as the next best heuristic, similarity to known persons \((n = 6)\). The remaining heuristics, face recognizability at presentation, sympathy or antipathy, and difficulty of recall were stated as useful five, four, and three times, respectively.

**Table 1**

| Degree of Usage of Various Heuristics in Making Recognition Predictions and Their Usefulness in the JOL Group |
|---|---|---|---|
| Heuristic | Usage | Very Frequent | Usefulness |
| | | | |
| Face recognizability at presentation | 1 | 6 | 3 | 2 | 0 | 5 |
| Difficulty of recall | 2 | 2 | 6 | 2 | 3 | 3 |
| Difficulty to memorize | 0 | 2 | 4 | 6 | 4 | 2 |
| Distinctive facial features | 0 | 0 | 2 | 10 | 10 | 0 |
| Sympathy or antipathy | 5 | 3 | 4 | 0 | 2 | 1 |
| Similarity to known persons | 3 | 6 | 1 | 2 | 4 | 2 |

Note—Entries are number of subjects giving the respective rating.
its contralateral homologue, T4, were replaced by the grand mean ERPs at these electrodes over the other 11 subjects of this group for each condition. ERPs were quantified by several average voltage measures related to a 100-msec prestimulus baseline period. Average amplitudes were calculated for consecutive 100-msec intervals from 200 to 600 msec and for the interval between 600 and 1,000 msec.

Amplitudes of the study-phase ERP were analyzed in three main steps. Step 1 tested for amplitude differences between subsequently recognized and unrecognized faces— that is, for the presence of a Dm. Step 2 tested for amplitude differences between JOLs and distinctiveness ratings. Finally, scalp topographies of the Dms and the rating effects were compared after correcting for overall amplitude differences. To reduce the risk of falsely rejecting the null hypothesis, the first analysis in each step was a global multivariate analysis of variance (MANOVA) of all time segments for an overall test of condition effects (Wilks's lambda test statistic). These analyses included a group factor, repeated measures on the experimental effect in question, and repeated measures on all 16 electrode sites, treated as one factor. Since the MANOVAs indicated in each case that the effects of interest were significant, separate Bonferroni-corrected ANOVAs of these effects were conducted for the individual time segments. In these cases, the Huynh-Feldt method was used to correct for any violations of the sphericity assumption. Please note that, due to the average reference derivation, for each ERP the amplitudes over the 16 electrodes add to zero. Therefore, condition effects in MANOVAs and ANOVAs are meaningful only in interaction with electrode site.

ERPs predictive of face recognition. Figure 1 presents the ERP waveshapes from the study phase, averaged according to whether or not the faces were subsequently recognized. It appears as if the ERPs were more positive for subsequently recognized faces than for unrecognized faces at the anterior electrodes and more negative at posterior and temporal sites.

The global MANOVA across all five time segments (Table 2) indicated a highly significant interaction of electrode site with subsequent recognition and, albeit considerably weaker, with group alone and with both subsequent recognition and group. The ANOVAs of the individual time segments (Table 3) confirmed that, following the 300- to 400-msec segment, there were increasingly significant interactions between electrode site and subsequent recognition, indicating amplitude differences dependent on memory (Dm). Between 300 and 1,000 msec, the Dm was also modulated by group. The group × electrode interaction in the MANOVA could not be localized within individual time segments.

ERPs according to JOLs and distinctiveness. Figure 2 depicts the ERPs from the study phase, averaged according to JOLs and distinctiveness ratings. The global MANOVA for the average amplitude measures of these ERPs across the five time segments (Table 4) included a group factor and repeated measures on rating (positive vs. negative JOLs in Group 1 and high vs. low distinctiveness for Group 2). In addition to electrode site, we also considered the difference between ERPs to contrast-manipulated and contrast-unmanipulated faces as a separate within-factor contrast. This factor tested whether any rating effects for contrast-unmanipulated faces, based on the same ERPs as those used for assessing Dm, appear also for the manipulated faces that had not been used for the Dm. The MANOVA indicated highly significant electrode × rating and electrode × contrast interactions. There were also weak interactions of electrode site with group and with rating, group, and contrast. The ANOVAs of individual time segments (Table 5) showed that the rating effects appeared during the same time segments where recognition effects had also been present—that is, from 300 to 1,000 msec.

It is important to note that these effects were not a mere artifact of sorting the ERPs according to a criterion that is correlated with recognition, because the recognition effects were not modulated by the contrast of the stimuli: the fourfold interaction that appeared in the MANOVA did not appear in any of the ANOVAs. Also,
there was no interaction of rating with group, indicating that JOLs and distinctiveness ratings yield very similar effects in both size and topography. According to the ANOVAs, the interaction between contrast and electrode is localized in the 500- to 1,000-msec time range.

**Topographical analyses of study-phase ERPs.**

Figure 3 shows the Dms and the rating effects as ERP difference waves. Figure 4 depicts maps of the respective topographies for those time segments where the effects had been significant. The Dms and the rating effects appear to show similar scalp distributions, particularly for the time segments between 300 and 500 msec, with more positivity over frontocentral sites and more negativity over occipitotemporal sites, whereas some topographic variability seems to exist in the later segments. However, there was also considerable amplitude variation between the difference waves and over time segments. Compare, for example, the large Dms and the small JOL effect. Amplitude differences given constant topography relate to the strength of generator activity in the brain. Source strength variations affect the ERP voltages at the different electrode sites in a multiplicative fashion. Because of the additive model underlying ANOVAs, interactions may arise between electrode and experimental conditions even when conditions differ only in source strength but not in generator location or orientation. When effects differ in overall amplitude, it is therefore unclear whether interactions with electrode site indicate true differences in topography and hence in generating brain systems or whether they merely arise because of the multiplicative nature of the source strength effect. To avoid this ambiguity, McCarthy and Wood (1985) suggested scaling amplitudes over electrodes prior to ANOVA (profile analysis). After scaling, interactions between experimental conditions and electrode site can be unambiguously interpreted as indicating true differences in generator source location or orientation.

The amplitude differences for each condition, time segment, and subject were scaled to the same amplitude range over the 16 electrodes and submitted to a single MANOVA. Apart from the different time segments

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrode</td>
<td>75,1671</td>
<td>13.8</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Group × electrode</td>
<td>80,1680</td>
<td>1.4</td>
<td>.017</td>
</tr>
<tr>
<td>Rating × electrode</td>
<td>80,1680</td>
<td>1.8</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rating × group × electrode</td>
<td>80,1680</td>
<td>0.8</td>
<td>.912</td>
</tr>
<tr>
<td>Contrast × electrode</td>
<td>80,1680</td>
<td>2.1</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Contrast × group × electrode</td>
<td>80,1680</td>
<td>0.9</td>
<td>.639</td>
</tr>
<tr>
<td>Contrast × rating × electrode</td>
<td>80,1680</td>
<td>1.0</td>
<td>.507</td>
</tr>
<tr>
<td>Rating × contrast × group × electrode</td>
<td>80,1680</td>
<td>1.5</td>
<td>.002</td>
</tr>
</tbody>
</table>
Table 5  
**F Values and Significance Levels of ANOVAs of Average ERP Amplitudes According to Distinctiveness and JOL Ratings**

<table>
<thead>
<tr>
<th>Source</th>
<th>Time Range (msec)</th>
<th>200–300</th>
<th>300–400</th>
<th>400–500</th>
<th>500–600</th>
<th>600–1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>F</em></td>
<td><em>MS</em></td>
<td><em>F</em></td>
<td><em>MS</em></td>
<td><em>F</em></td>
<td><em>MS</em></td>
</tr>
<tr>
<td>Electrode</td>
<td>11.2§</td>
<td>2,401.5</td>
<td>14.3§</td>
<td>3,102.0</td>
<td>28.5§</td>
<td>4,204.0</td>
</tr>
<tr>
<td>Group × electrode</td>
<td>2.2</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Contrast × electrode</td>
<td>0.9</td>
<td>92.0</td>
<td>1.4</td>
<td>103.6</td>
<td>0.4</td>
<td>117.7</td>
</tr>
<tr>
<td>Contrast × group × electrode</td>
<td>1.6</td>
<td>1.8</td>
<td>1.7</td>
<td>1.0</td>
<td>2.5†</td>
<td></td>
</tr>
<tr>
<td>Rating × electrode</td>
<td>0.9</td>
<td>79.0</td>
<td>3.5†</td>
<td>128.4</td>
<td>4.0‡</td>
<td>150.0</td>
</tr>
<tr>
<td>Group × rating × electrode</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Contrast × rating × electrode</td>
<td>1.0</td>
<td>62.9</td>
<td>1.2</td>
<td>63.4</td>
<td>0.8</td>
<td>76.3</td>
</tr>
<tr>
<td>Group × contrast × rating × electrode</td>
<td>0.9</td>
<td>0.7</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td></td>
</tr>
</tbody>
</table>

Note—df = 15,330.  *p = .05.  †p < .05.  ‡p < .01.  §p = .001.

(300–1,000 msec), the MANOVA involved a group factor, as well as repeated measures for electrode site and the type of difference wave. The latter factor included levels Dm (subsequently recognized minus unrecognized), rating effects in contrast-unmanipulated faces [positive minus negative JOL (Group 1), high minus low distinctiveness (Group 2)], and rating effects in contrast-manipulated faces. This MANOVA revealed a difference wave × electrode interaction, which indicates topographical differences between the Dm and the rating effects for at least one time segment (Table 6).

To pinpoint the topographical difference, we calculated ANOVAs of the scaled amplitudes with the same factors as those used for the MANOVA at each individual time segment. These ANOVAs did not show any significant effects (ps > .20) for the time segments 300–400 msec and 400–500 msec. In contrast, type of rating had effects in both subsequent time segments [F(1,306) = 1.8 and 2.1, MS = 0.058 and 0.055, ps < .05 and .01, respectively]. This result indicates that the topographies of the Dm and the rating effects are similar between 300 and 500 msec but differ between 500 and 1,000 msec. Al-

---

Figure 3. Difference waves between ERPs to subsequently recognized and unrecognized faces for each subject group superimposed with differences waves between ERPs to faces with positive and negative JOLs (left panel) and between distinctive and typical faces (right panel).
Figure 4. Scalp topographies of the differences in average ERP amplitude for the 400- to 500-msec time segment. The top panels show the differences between subsequently recognized and unrecognized faces (Dm) for each subject group. The two lower panel rows show amplitude differences between ERPs with good and poor JOLs of Group 1 (left panel) and between high- and low-distinctive faces (right panel).
Table 6
Global Topographic Comparison Between ERP Differences According to Subsequent Recognition (Dm) and According to JOLs and Distinctiveness Ratings with MANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrode</td>
<td>60,1365</td>
<td>2.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Group × electrode</td>
<td>64,1369</td>
<td>1.3</td>
<td>.074</td>
</tr>
<tr>
<td>Difference × electrode</td>
<td>128,2791</td>
<td>1.7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Difference × group × electrode</td>
<td>128,2791</td>
<td>0.8</td>
<td>.914</td>
</tr>
</tbody>
</table>

though it appears unlikely that the Dms or the rating effects are due to ocular artifacts, the ocular activity shown by Group 1 in the later part of the recording epoch may have affected the scalp topographies of the reported effects. Therefore, we refrained from a detailed post hoc analysis of group-related topographical effects during the 500- to 1,000-msec time segments.

Recognition phase ERPs. For the ERPs from the recognition phase, average amplitude measures were calculated for the same time windows as those for the study-phase ERPs. ANOVAs for the individual time segments did not show any significant rating-related effects.

DISCUSSION

This study demonstrates that subjects can predict their own recognition performance for faces. The recognition ratings not only discriminated between old and new faces, they also correlated with the preceding predictions for recognition (JOLs). This conforms with the report by Vokey and Read (1992) that rated memorability of faces predicts recognition in a different subject group. The magnitude of the correlation between JOLs and recognition is in the same order as commonly reported for verbal material.

Behavioral evidence supports the hypothesis that, in making JOLs, the subjects rely at least partially on facial distinctiveness. Distinctiveness was the strategy reported most frequently postexperimentally, and distinctiveness ratings predicted recognition performance about as efficiently as the JOLs. There was also a sizeable correlation over faces of facial distinctiveness and JOLs, rated in different subject groups, conforming with a similar correlation within the same subject group reported by Vokey and Read (1992). The relationship between distinctiveness and recognition was unaffected when JOLs were partialled out, but the relationship between JOLs and recognition was reduced to insignificance by controlling for distinctiveness. From this pattern, one may conclude that the validity of JOLs for face recognition depends on variations of facial distinctiveness. The perceived contrast of the faces, on the other hand, appears to have played some role for JOLs and distinctiveness ratings, with clear faces being rated as more distinctive and yielding higher JOLs than flat faces. However, contrast was not functional for the predictiveness of these ratings for recognition, because contrast was unrelated to recognition and the relationships between JOLs and distinctiveness with recognition were unaltered when contrast ratings were partialled out.

Therefore, a major and valid heuristic of the subjects in predicting face recognition appears to be related to their (correct) metacognitions that distinctive faces are more likely to be recognized. Although this seems to be the natural interpretation, it should be noted that the causal relationship may also be the other way around. It is conceivable that subjects directly or indirectly monitor the quality of the memory trace and base distinctiveness ratings on monitoring outcome. On the other hand, the latter suggestion is difficult to reconcile with the robustness of the relationship between JOLs and distinctiveness if trace strength, as measured by recognition ratings, is partialled out.

The present study replicates our previous report (Sommer et al., 1991) of an ERP component (Dm) that predicts subsequent recognition of faces. Because there was no correlation between contrast and recognition ratings, Dm appears to be unrelated to perceptual performance or fluctuations of attention during study. Since the scalp topographies of the Dm and the basic waveforms differed over the entire 700-msec epoch where it was present (see Note 4), the Dm is not just a modulation of the basic ERP waveform observed during the respective time segment. Rather, the Dm must be considered as a component with a time course and brain source separate from those of the various components in the base waves. The Dm, as observed in our previous study and the present study, did not begin before 200 and 300 msec after stimulus onset, respectively, and lasted for several hundred milliseconds. Thus, the cognitive processes related to Dm appear to be of a rather long-lasting type. In addition, the onset of the Dm lies beyond the 200 or so milliseconds following stimulus onset during which ERP components are observed that directly reflect perceptual processes. Even structural encoding (Bruce & Young, 1986) of face-specific features appears to start well before our Dm onset; Oram and Perrett (1992) have shown that neurons in monkey brain discriminate between different head views already after 25 msec. Although it may be risky to argue on the absence of an early ERP effect, the late onset of the Dm and the long duration may indicate that it does not directly relate to perceptual or structural encoding of faces.

As in our previous study, there was little difference during the recognition phase between ERPs to recognized and unrecognized old faces, the same stimuli that had elicited the Dm when presented during study. Therefore, it is unlikely that the Dm is related to preexisting differences between faces, for example, in attractiveness, emotional connotations, or a priori familiarity. These results are compatible with the suggestion of Sommer et al. (1991) that our Dm is a manifestation of processes that relate to the formation of representations for faces in memory.

What can be concluded from the ERPs for JOL- and distinctiveness-related cognitive processes? From the time course of these effects in the ERPs, we may again
infer that the processes are long-lasting and resemble late-onset processes rather than early onset processes. This contrasts to Reder's (1987, 1988) suggestion of a "rapid feeling of knowing" and may provide support for Leonesio and Nelson's (1990) results, which indicate that FOKs and JOLs may be based on different underlying structures.6

The similarity of the JOL and distinctiveness effects in ERPs to the Dm in terms of both time course and scalp topography argues for a common basis of these phenomena. It appears that the cognitive processes underlying subsequent recognition performance, recognition predictions, and distinctiveness ratings relate to the activity in similar brain systems. If we infer from similar brain processes to similar cognitive processes, it may be suggested that the behavioral correlation between distinctiveness ratings and JOLs in the present study or between distinctiveness and normative memorability ratings (Vokey & Reed, 1992) emerges because they are based on similar processes.

As to the nature of these processes, at least two suggestions can be made. First, it is conceivable that facial distinctiveness is the basic variable. If distinctive faces initiate quantitatively or qualitatively different ERP-related processes than do typical faces and if distinctive faces are recognized more readily, then we might expect similar ERP effects also on the basis of recognition differences. Alternatively, if our Dm reflects engram formation, as suggested above, the basic process may be the formation of new representations for faces in memory that happens to be more intense in distinctive faces. It may not be possible to distinguish between these alternative views at present, particularly considering that, on the one hand, the behavioral results above suggested that distinctiveness rather than face strength is the basic variable. On the other hand, if distinctiveness is more basic than face strength, one should expect, contrary to what we observed, Dm-like ERP differences also during the recognition phase when ERPs are sorted according to old/new ratings, because the respective ERP-eliciting faces differ also in distinctiveness.

In sum, subjects can predict their own recognition performance for faces. These predictions appear to be based at least to some extent on the perceived distinctiveness of the face. Recognition predictions, facial distinctiveness, and later recognition are all linked to ERP differences that start relatively late and are indistinguishable in scalp topography, consistent with the possibility of a common basis at the level of underlying brain processes.

REFERENCES


NOTES

1. The widespread spatial distribution, often reported for Dms, may relate to the choice of the reference electrode. Most studies used reference electrodes at or close to the mastoids, which are adjacent to anterior temporal lobe structures. However, as memory functions have been linked to the temporal lobes of the brain (Squire, 1987), these structures might well contribute to the neurogenesis of Dm. When a broadly distributed scalp-recorded Dm is found with a common reference electrode, it is impossible to distinguish widespread brain activity from local activity close to the reference electrode that is subtracted from all other EEG channels. In our previous study and in the present study, we recorded from a large array of electrodes and used the average activity of all channels as reference (average reference). This makes the recording independent from the reference site and enhances topographical differences by eliminating the activity shared by all electrodes (Lehmann, 1987).

2. In the signal detection literature, it is sometimes recommended to use at least six rating categories for determining P(A). To assess the risk involved in using only four categories, we ran several simulations with the same measurement procedure (trapez formula) and comparable number of ratings and subjects as in the present study. Apart from the expected about 10% underestimation of P(A) with four as opposed to six rating categories, the power for detecting above-chance sensitivities at α = .01 was better than 99.5% even for P(A) as low as about .6.

3. In Figures 1 and 2, systematic electrooculographic activity is evident, particularly for Group 1. This is probably related to the tendency of some subjects to direct their gaze toward the eye and hairline region of the face stimuli. Even more disturbing than the EOG activity per se is a condition-related EOG difference, particularly in Group 1. Making the artifact rejection criterion more conservative did not eliminate these EOG effects leading to an unacceptable low number of trials per average in many subjects. Therefore, it is legitimate to ask whether the condition effects in the ERPs may derive from systematic differences in the eye movements. This argument can be rejected on several grounds. For one, there is little interaction of the relevant effects with the groups, even though the systematic eye movements are notably different between groups. Second, the distribution of the ERP effects is incompatible with what would be expected for eye-movement artifacts. While the latter should yield the strongest effects at prefrontal and frontal leads, leading out at posterior sites, the experimental effects observed are not disproportionately strong at Fpz. Third, we empirically tested this conjecture for those 8 subjects (4 from each group) who had the strongest systematic eye movements and also sizeable Dms. We selected and averaged trials with pronounced systematic eye movements (EOG ranges from 40 to 80 μV, but no blinks) and measured the average ERP activity in the time segment 600-1000 msec—the time during which eye movements are maximal. We then compared the topographies of this eye-movement-affected ERP activity and of the Dm for the same time segment by means of profile analyses, reasoning that the distributions should be similar if the Dm is due to systematic eye movements. However, as it turned out, the topographies were significantly different [F(15, 105) = 2.65, p < .05]. Therefore, we can reasonably exclude that the Dm—or the JOL or distinctiveness effects that resemble the Dm—are caused by systematic electrooculographic contributions.

4. An interesting issue for the interpretation of the Dm is whether it constitutes a modulation of the basis ERP waveform(s) or whether it can be considered a separate component. Profile comparisons were conducted for the individual time segments 300-1000 msec between the Dm and both its basic waveforms—that is, between ERPs to subsequently recognized faces, unrecognized faces, and the difference between them (the Dm) as three levels of one factor. This analysis revealed highly significant differences in topography for all four segments (all ps < .001), which were also significant in all pair-wise comparisons between the Dm and each of the two basic waveforms.

5. There are several studies of the von Restorff effect, which may be considered as a variant of distinctiveness, and the P300 component of the ERP (e.g., Fabiani, Karis, & Donchin, 1990; Karis, Fabiani, & Donchin, 1984). However, we will not discuss these studies here because their relevance for facial distinctiveness is unclear due to conceptual and procedural differences. May it suffice to say that our distinctiveness effect is unrelated to the P300 component because its scalp topography is very different from the parietocentral maximum typically observed for P300 (cf. Figure 4).

6. We are grateful to Tom Nelson for this suggestion.

(Manuscript received April 30, 1993; revision accepted for publication February 16, 1994.)