Encoding Activity and Face Recognition
Christian Coin and Guy Tiberghien
Institut de Sciences Cognitives, Lyon, France

A series of studies conducted over the past 20 years have explored the effects of various tasks on recognition memory for faces. Memory for faces appears better when the study task involves judgements about an abstract trait rather than a physical feature. The various situations in which these results were obtained raise important methodological questions regarding the learning conditions, whether incidental or intentional, and the duration of exposure to the stimulus during the study phase. We consider here two alternative explanations for the reported results. One concerns depth of processing and the other the opposition between component and holistic processing. Possible strategies for improving face recognition performance are considered.

INTRODUCTION

We can read expression from a face presented within 20ms (Simpson & Crandall, 1972) and recognise former schoolmates at about a 90% recognition rate 35 years after graduation (Bahrick, Bahrick, & Wittlinger, 1975). In the past 20 years, numerous studies have explored the influence of encoding instructions on facial recognition. Most of the studies examined here, including the most recent (e.g. Terry, 1993), used an experimental paradigm similar to that of the pioneering study reported by Bower and Karlin (1974). Studies employing adults (e.g. Bower & Karlin, 1974; Clifford & Prior, 1980; McKelvie, 1991; Patterson & Baddeley, 1977; Sporer, 1991; Wells & Turtle, 1988; Winograd, 1981) show that the requirement for personality trait judgements, such as likeability, yield higher recognition rates that judgement about size of a feature or gender. Similar patterns of results have been obtained with children (7 to 14 years old: Carey, Diamond, & Woods, 1980), with older subjects (50 to 70 years...
old: Warrington & Ackroyd, 1975) and with amnesic Korsakoff patients (Biber et al., 1981). This paper will review and evaluate the various procedures employed in these studies.

There are four categories of encoding instructions for the study phase of the experiments reviewed here:

1. Standard instructions in which subjects are told only that a retention test follows the study phase.
2. Instructions to attend to physical characteristics viewed directly (e.g. facial features and accessories like glasses), or inferred (e.g. gender, race, weight and height).
3. Instructions to regard global aspects of the face (e.g. expression, shape), personality traits (e.g. likeability and intelligence). These instructions are sometimes related to identifying activities of the figure (e.g. profession, hobby), or to religious or political preference.
4. Instructions to choose the most distinctive facial feature among a set of facial features.

Taken together, the results show that judgements of personality traits yield better recognition performance than those of physical traits. The data provided in Table 1 also demonstrate that inferential judgements improve distinguishability (d’) among them. The decision criterion (β), appears unaffected by the encoding conditions (see also Shapiro & Penrod, 1986) although the data supporting this conclusion is sparse. The faces searched for their most distinctive physical feature among a set designated by the experimenter were better remembered than those judged with respect to a single physical feature (Courtois & Mueller, 1979; Sporer, 1991; Winograd, 1981). Finally, recognition performance was not different following judgements about personality traits than after judgements about the most distinctive facial feature (Daw & Parkin, 1981; Deffenbacher, Leu, & Brown, 1981; Parkin & Goodwin, 1983; Parkin & Hayward, 1983; Winograd, 1981).

METHODOLOGICAL ASPECTS

Before attempting an examination of the proposed explanations for these results it is important to note the differences in the methodologies that underlie them. In some procedures subjects were told that a recognition test would follow the study phase. In other experiments this information was not given. Additionally, exposure duration of the target face varies among the studies. Possible consequences of such variations must be analysed. In particular, the subjects could adopt a particular encoding strategy when they know that a test would follow. Alternatively their normal strategy might be unaltered by the instructions. Finally, if different judgements have different processing times
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Hits</th>
<th>FA</th>
<th>$d'$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>$P$</td>
<td>$F$</td>
<td>$P$</td>
</tr>
<tr>
<td>Biber et al. (1981)</td>
<td>.65 &lt; .81</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bloom &amp; Mudd (1991)</td>
<td>.56 &lt; .76</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bower &amp; Karlin (1974; Exp I)</td>
<td>.60 &lt; .81</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bower &amp; Karlin (1974; Exp II)</td>
<td>.56 &lt; .76</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Carey et al. (1980)$^e$</td>
<td>.80 &lt; .88</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Courtois &amp; Mueller (1979)</td>
<td>.69 &lt; .84</td>
<td>21 &gt; 12</td>
<td>1.5 &lt; 2.4</td>
<td>2.1 – 2.7</td>
</tr>
<tr>
<td>Devine &amp; Malpass (1985)</td>
<td>—</td>
<td>—</td>
<td>.86 &lt; .80</td>
<td>—</td>
</tr>
<tr>
<td>Gaonac'h &amp; Giboin (1977; Exp I)</td>
<td>.85 &lt; .92$^c$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gaonac'h &amp; Giboin (1977; Exp II)</td>
<td>.78 &lt; .87$^c$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Light et al. (1979; Exp I)$^d$</td>
<td>.54 &lt; .65</td>
<td>.29 &gt; .21</td>
<td>.66 &lt; .75$^e$</td>
<td>—</td>
</tr>
<tr>
<td>Light et al. (1979; Exp II)$^d$</td>
<td>.55 &lt; .64</td>
<td>.28 &gt; .21</td>
<td>.66 &lt; .76$^e$</td>
<td>—</td>
</tr>
<tr>
<td>Mayes et al. (1980)$^g$</td>
<td>.42 &lt; .50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>McKelvie (1978)</td>
<td>.83 &lt; .88</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>McKelvie (1985)</td>
<td>.82 &lt; .87</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>McKelvie (1991)</td>
<td>.66 &lt; .90</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mueller et al. (1978; Exp I)$^f$</td>
<td>.58 &lt; .66</td>
<td>.20 = .19</td>
<td>1.1 &lt; 1.4</td>
<td>1.7 = 1.4</td>
</tr>
<tr>
<td>Mueller et al. (1978; Exp II)</td>
<td>.48 &lt; .67</td>
<td>.25 = .23</td>
<td>.69 &lt; 1.2</td>
<td>1.5 = 1.2</td>
</tr>
<tr>
<td>Mueller et al. (1979)</td>
<td>.63 = .69</td>
<td>.22 &gt; .15</td>
<td>1.2 &lt; 1.7</td>
<td>1.8 = 2.7</td>
</tr>
<tr>
<td>Mueller &amp; Wherry (1980)</td>
<td>.52 &lt; .78</td>
<td>.16 = .15</td>
<td>1.2 – 1.9</td>
<td>3.0 – 1.5</td>
</tr>
<tr>
<td>Patterson &amp; Baddeley (1977)</td>
<td>.76 &lt; .82$^f$</td>
<td>—</td>
<td>1.5 &lt; 1.9</td>
<td>—</td>
</tr>
<tr>
<td>Smith &amp; Winograd (1978)$^h$</td>
<td>.75 &lt; .86</td>
<td>.21 = .22</td>
<td>1.5 &lt; 1.9</td>
<td>1.0 = 0.9</td>
</tr>
<tr>
<td>Sporer (1991)</td>
<td>.56 &lt; .66</td>
<td>.30 – .26</td>
<td>.69 &lt; .77$^b$</td>
<td>—</td>
</tr>
<tr>
<td>Stringer (1987)</td>
<td>.51 &lt; .67</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Strnad &amp; Mueller (1977)$^i$</td>
<td>.46 &lt; .55</td>
<td>.26 = .34</td>
<td>—    + —</td>
<td></td>
</tr>
<tr>
<td>Terry (1993; Exp I)</td>
<td>.74 &lt; .81</td>
<td>.16 = .16</td>
<td>2.4 = 2.7</td>
<td>5.1 = 4.4</td>
</tr>
<tr>
<td>Terry (1993; Exp II)</td>
<td>.65 &lt; .70</td>
<td>.24 &gt; .20</td>
<td>1.4 &gt; 1.8</td>
<td>3.0 = 3.3</td>
</tr>
<tr>
<td>Warrington &amp; Ackroyd (1975)</td>
<td>.80 &lt; .86$^c$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wells &amp; Hryciw (1984)</td>
<td>.25 &lt; .55</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wells &amp; Turtle (1988)</td>
<td>.54 &lt; .78</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Winograd (1976)</td>
<td>.66 &lt; .75$^c$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Winograd (1981; Exp I)</td>
<td>.64 &lt; .73</td>
<td>.25 &gt; .16</td>
<td>1.0 &lt; 1.6$^e$</td>
<td>—</td>
</tr>
<tr>
<td>Winograd (1981; Exp II)</td>
<td>.70 &lt; .76</td>
<td>—</td>
<td>.85 &lt; .88$^c$</td>
<td>—</td>
</tr>
</tbody>
</table>

$F$ = judgement about a facial feature; $P$ = judgement about personality trait. Dashes indicate no data available. The differences are significant at the .05 level at least; < inferior; > superior; = no significant difference; – no difference mentioned.

$^e$Data estimated as accurately as possible from a figure in the original article. $^f$The sensitivity is expressed by the $A'$ estimate. $^g$Hits plus False Positives. $^h$Data obtained from typical and unusual faces. $^i$The sensitivity is expressed by the $A_g$ estimate. $^j$Data obtained from two groups of subjects (5 and 15 seconds exposure duration). $^k$Significant at the .10 level. $^l$Data obtained from two groups of subjects (18–52 years old and 50–80 years old). $^m$Data obtained from two groups of subjects (incidental and intentional learning).
and exposure duration is constant, the effects of exposure duration may be confounded with the type of encoding instruction.

Control Group, Incidental Learning, and Intentional Learning

Studies designed to examine the effects of encoding instructions on facial recognition vary in the judgements required. This has two important consequences. First, they relate only to a limited set of strategies for face encoding employed outside laboratory settings. Normally face encoding is conducted spontaneously and is free of imposed judgements (at least of the kind discussed here). Second, comparison between studies is difficult because of the lack of a common basis, in addition to unavoidably varying procedures between experiments. For example, it is difficult to draw a parallel between recognition performance obtained after judgements about sex and about likeability (e.g. Bower & Karlin, 1974), and those obtained after judgements about weight and intelligence (e.g. Mueller, Bailis, & Goldstein, 1979).

In these studies, three types of learning were used:

1. Standard intentional learning in which subjects are instructed only to examine the stimuli during the study phase. They are warned about the subsequent test phase.
2. Non-standard intentional learning in which subjects make judgements during the study phase. They are warned about the subsequent test phase.
3. Incidental learning in which subjects make judgements without knowing about the subsequent recognition test.

In order to assess the effects of different encoding instructions with regard to the standard learning paradigm (basic level), the learning factor must not be confounded with the type of judgement factor made during the inspection phase (e.g. Devine & Malpass, 1985; Sporer, 1991). In these experiments the learning during the inspection phase is intentional for one group, and incidental with judgements to be made for the other group. Thus, it is not possible to compare the effect of standard intentional learning recognition with the effect of intentional learning with judgements. However, Bower and Karlin (1974) showed that facial recognition was superior following judgements about the honesty of a face than after judgements about its gender, both in an incidental learning condition (Experiment 1; see also Carey et al., 1980; Patterson &

---

1The basic level considered here is the condition where the subject does not make judgements during the study phase and is notified of the subsequent test phase. Perhaps it would be preferable for the basic level to be the condition of incidental learning in which no judgements are made. However, it must be admitted that it is difficult to create such a condition.
Baddeley, 1977; Terry, 1993) and in an intentional learning condition (Experiment 2; see Sporer, 1991; Wells & Turtle, 1988). The same pattern of results was obtained by Strnad and Mueller (1977) where incidental and intentional learning conditions were combined within the same experiment. Judgements based on the most distinctive feature do not lead to better recognition performance than judgements about the inferred character of the face in the incidental learning condition (Daw & Parkin, 1981; Valentine & Bruce, 1986), or in the intentional learning condition (Winograd, 1981). In conclusion, the similarity of the results obtained with intentional and incidental conditions seems clearly demonstrated. To notify subjects of the subsequent mnesic test does not have an effect on recognition performance.

On the other hand, informing subjects about the recognition test could yield an effect of encoding instructions that is hard to interpret. We cannot be absolutely sure that subjects did not use their usual encoding strategy whatever the instructions given. In the same way that subjects cannot help semantically identifying a written word (Stroop effect, Stroop, 1935), or that one has difficulty hearing speech as simple sounds (Fodor, 1983), asking subjects to make a judgement about a facial feature does not assure us that a judgement about personality traits is not also made. To ensure that subjects actually used an holistic encoding strategy, Deffenbacher et al. (1981) asked them to give an opinion about the difficulty of memorising the overall structure of the face. Similarly, in other studies, instead of a binary judgement (e.g. likeable vs. unlikeable), subjects were asked to make their judgement on a 4-point scale (Biber et al., 1981) or on a 7-point scale (Wells & Hryciw, 1984). In this way one can hope to have a more effective contribution by subjects. To say whether a nose is short or long could hardly hold the subjects’ attention. In order to offset this disadvantage, Sporer (1991) alternated within a session the feature on which judgements were made. For example, subjects were asked about nose size at item 1, about the gap between the eyes at item 2, etc. However these precautions are not entirely satisfying. The most direct control was obtained by Bloom and Mudd (1991). They measured together time inspection, eye movements and pupil diameter. They found that as more eye movements were recorded, more features were processed. However this recording is useful only if we consider, as Bloom and Mudd did, that personality trait judgements require the examination of numerous physical features and cannot be done at first glance. In summary, it is difficult to obtain an equivalence between the encoding requested and the encoding made.

Taking into account studies where no judgements were made during encoding, the data show consistently that memory for faces is invariably superior or equally following no judgements (standard intentional encoding) compared with following judgements about a facial feature (see Table 2). Moreover, facial recognition is superior or equal following judgements about a personality trait compared with recognition after no judgements, and after a judgement about a facial feature. So we could state that judgements about a
personality trait of a face lead to better recognition performance than no judgement (standard intentional encoding) which in turn leads to a better recognition performance than judgements about a physical feature. Therefore, encoding instructions involving judgements about a face facilitate (when a personality trait judgement is requested) or disrupt (when a facial feature judgement is requested) the recognition of the face.

### Exposure Duration

Several authors argue that judgements about personality traits require more time to be achieved than do judgements about facial features (e.g. Baddeley & Woodhead, 1982; Daw & Parkin, 1981). Goldstein and Papageorge (1980) obtained data that seems to invalidate such a hypothesis. Judgements about beauty and attractiveness of the face were done within 150ms, which is much shorter than the time allowed to subjects to make judgements in the present review (i.e. 5 to 15 seconds). In contrast, Bloom and Mudd (1991, Experiment 1) recorded inspection time where subjects were asked to close their eyes as soon as they had made their judgement. The results showed that no judgement and gender judgement did not require different processing time (0.39 s vs. 1.09 s, no significant difference), but that these both required less time than honesty judgements (2.78 s).

The exposure duration factor could be confounded with the encoding type factor. In Montgomery’s study (cited in Deffenbacher et al., 1981) the recognition performance varied with encoding conditions made within eight
seconds. On the other hand, no recognition performance difference occurred between encoding conditions when the presentation of the target stimulus was repeated. The ceiling effect that could appear with a long exposure duration does not allow clear interpretations. Only experiments in which both encoding instructions and duration of study process vary could allow us to dissociate the effect of these two factors. For example, Daw and Parkin (1981; see also Parkin & Hayward, 1983, Experiments 1 and 4) recorded the decision time required to make varied judgements about faces. The results show that judgements about personality traits are made more quickly than judgements about a distinctive facial feature (2255 ms vs. 3280 ms). If we consider the results obtained by Bloom and Mudd (1991) and by Parkin and his collaborators (about a physical feature, a personality trait, and a distinctive facial feature), any judgements could be achieved within three seconds. Unfortunately, in the studies examined here, the encoding time was always greater than five seconds, so it is not possible to dissociate the exposure duration factor and the encoding type factor. A more satisfactory paradigm would allow subjects only the duration necessary to make a judgement. Indeed, it could be argued that in the facial feature judgement condition, as the duration allowed is greater than the necessary duration to make such a judgement, subjects could also have sufficient time to make a personality trait judgement.

Exposure Duration and Judgement about a Distinctive Feature

In the distinctive feature judgement condition (e.g. Reynolds & Pezdek, 1992, Experiment 2; Valentine & Bruce, 1986, Experiment 3; Winograd, 1978, 1981) the subjects were asked to examine the facial features and to select the feature that seemed to be the most distinctive. The level of recognition performance obtained was not different from the recognition rate obtained following a personality trait judgement. The authors consider that this task requires the extraction of a large number of features (e.g. Winograd, 1981). However the recognition performance is rather varied, that is, from 68.8% to 91.2%. In all these experiments, the encoding instruction factor is a between-subjects factor. So, each subject was asked to perform only one task, which consisted of either making a judgement about a personality trait or choosing the most distinctive feature. Moreover, the list from which the subjects selected a feature did not vary within the study, and this list contained a various number of features (from 5 to 11). The subjects knew which features they had to observe during the exposure of the target face and it was probably easier to observe five features during five seconds (Courtois & Mueller, 1979) than nine features during three seconds (Daw & Parkin, 1981). So it can be assumed that the performance obtained depended in part on the exposure duration of the target face (from 3 to 20 seconds), and also depended on the
length of the list from which a feature was selected (from 5 to 11) (see Fig. 1).

**INTERPRETATIONS**

**Depth of Processing**

The data showing that judging the inferred character of a face (e.g. likeability) leads to better recognition performance than a judgement about a physical feature, have been interpreted by others within the levels of processing framework (e.g. Bower & Karlin, 1974; McKelvie, 1985; Patterson & Baddeley,
1977; Winograd, 1976). The levels of processing hypothesis was demonstrated using verbal materials (Craik & Lockhart, 1972). These authors showed that words processed semantically were remembered better than words processed for their physical characteristics. Bower and Karlin (1974) were the first to apply this theoretical notion to memory for faces. The subjects were asked to make dichotomous judgements about gender and about the perceived honesty or likeability of a face. Bower and Karlin showed that memory for faces is considerably better following a deep encoding (e.g. judgement about likeability) than a superficial encoding (e.g. judgement about gender). They argue that as processing level increases, a larger number of associations to the target are formed in memory which enhances memory performance.

However, this hypothesis has been challenged by numerous authors. Winograd (1981) argues that it is the amount rather than the type of information encoded that enhances the recognition rate (elaboration hypothesis). Indeed, when the subjects were asked to choose the most distinctive facial feature, recognition performance was as high as following a judgement about a personality trait (Blaney & Winograd, 1978; Courtois & Mueller, 1979; Daw & Parkin, 1981; Deffenbacher et al., 1981; Parkin & Goodwin, 1983; Parkin & Hayward, 1983; Smith & Winograd, 1978; Valentine & Bruce, 1986; Winograd, 1978, 1981). More precisely, Winograd and his collaborators argue that a personality trait judgement involves the encoding of a large number of features, which enhances the probability of encoding a distinctive feature. Courtois and Mueller's (1979) hypothesis, although slightly different, leads to the same conclusion. According to these authors, a judgement about a personality trait (e.g. honesty) requires the examination of numerous facial features in order to allow the matching between the presented face and a memorised 'honest' face prototype. Thus, contrary to Winograd's (1981) ideas, the scanning of numerous features is required before judgement can be made. On the other hand, a judgement about a physical facial feature requires a decision on only one feature.

Kerr and Winograd (1982) criticise the inferential nature of such a hypothesis. Indeed, we cannot be sure that when the subjects make judgements about personality they ipso facto encode a large number of features. Kerr and Winograd found that face recognition increased with the number of descriptive sentences read during the encoding phase. Such a result is in agreement with Winograd's hypothesis. Most interestingly, in a third experiment they included a recall test to be sure that the encoding of the sentences was actually done by the subjects. Here the recognition rate did not vary with the number of sentences read during the encoding phase. Similarly, the recognition performance does not vary with the number of statements (3 vs. 15) shown with the target face (Baddeley & Woodhead, 1982). As suggested by an anonymous reviewer, an alternative argument could be proposed. That is, a longer than necessary exposure duration for the 'encode a feature/encode a distinctive feature'
instruction may allow subjects to encode more features than the instruction dictates is necessary, and this in itself may facilitate the involuntary judgement of personality.

If a judgement about personality induces the encoding of several facial features, a reconstruction task with an Identikit\(^2\) carried out after such a judgement would be done more successfully than following a physical feature judgement. Using a between-subjects factorial design, Wells and Hryciw (1984) asked their subjects to make a judgement about a personality trait or about a physical feature, and then submitted them to two different tests: either a recognition test or a reconstruction test. Although the recognition rate obtained confirms the earlier data, performance on the reconstruction task was better following judgements related to physical features than after judgements about the perceived honesty of the face. This result calls into question Winograd’s elaboration hypothesis. However we probably need to clarify the conditions in which such results have been obtained. First, the data that challenged Winograd’s hypothesis came from recall tests (Kerr & Winograd, 1982) and from reconstruction tasks (Wells & Hryciw, 1984). It can be assumed that such tasks require a different process from a recognition task (e.g. Comish, 1987). In a similar way, Baddeley and Woodhead (1982) did not observe any difference in recognition rate after the encoding of various statements presented with the target face. But this situation is rather different from the condition in which subjects are asked to make judgements about physical features.

Another attempt to test these two hypotheses was carried out by Bloom and Mudd (1991). In order to assess how many features of the target were examined in the two contrasted conditions (personality and feature judgements) they recorded the number of eye movements during the encoding phase. They observed that more eye movements were made in the personality trait judgement than in the feature judgement condition. However, although this result is compatible with the elaboration hypothesis (more features are encoded during personality judgements by more eye movements), it does not ipso facto infirm the depth of processing hypothesis.

**Level of Processing and Implication of the Subject**

Mueller et al. (1979) argued that a decision involving a self-reference criterion is deeper than one involving a neutral absolute criterion. In order to test this hypothesis, subjects were asked to base their decision on a personal criterion (e.g. ‘‘Is the person as tall as you?’’), or on an absolute criterion (e.g. ‘‘Is the person over 5ft 6in tall?’’). However, no significant difference was found between the recognition decisions made, whatever the type of encoding

---

\(^2\) An Identikit enables the composition or the reconstruction of a face from black and white photographs of the five main facial features; hair, eyes, nose, mouth, and chin.
instructions (see also Mueller, Courtois, & Bailis, 1981). It seems that the implication of the subject leads to a higher level of recognition. For example, when the question: “Does this person look like you?” is asked during the encoding phase, the recognition rate is superior to that following feature judgements but does not differ from those following personality judgements (Courtois & Mueller, 1979). In a rather similar way, Sporer (1991) obtained a better recognition rate when the subjects themselves generated relevant features for the judgement about a physical feature than after a judgement about a feature given by the experimenter. However, this effect was not observed with judgements about personality traits. Therefore, we may expect an interaction between the type of encoding and the implication of the subject.

**Componential and Holistic Processing**

Numerous authors have interpreted recognition performance as a function of the orienting tasks within the componential/holistic framework. Several types of judgements are seen as being holistic:

2. Judgements about a hobby or about the profession that the person depicted was likely to enjoy or to practise (Klatzky, Martin, & Kane, 1982; Sporer, 1991).
3. Judgements about the weight of the person depicted (Woodhead, Baddeley, & Simmonds, 1979).
4. Judgements about the global structure of the face (Deffenbacher et al., 1981).

On the other hand, judgements about physical facial features and about gender are supposed to be done in an analytical way (Galper & Costa, 1980; Wells & Hryciw, 1984). The componential/holistic dichotomy seems to be only a framework to interpret the results obtained. Indeed, none of the studies has directly compared the effects of holistic judgements (e.g., scanning the overall structure of the face) and componential judgements (e.g., looking at a facial feature) on subsequent recognition. Deffenbacher et al. (1981) asked subjects to perform an holistic task, i.e., to scan the global structure of a face. No difference was obtained in a yes/no recognition task whatever the encoding task (holistic, standard, or judgements about a distinctive feature).

*Judgements About Personality Trait or Intelligence.* If such a judgement could be carried out from one or two features, it could be performed analytically. On the other hand, the processing of numerous facial features could be either holistic or analytic. In a review about social attribution, Shepherd (1989) points
out that there is no real agreement among subjects about physical cues in ranking faces on intelligence. As various and different features are considered by subjects, it is likely that numerous physical facial cues are taken into account instead of a few, for example, thin eyebrows, large or small eyes. As such, we cannot know if the processing performance is analytic or holistic in nature.

**Judgements About Weight and Height.** Facial recognition rate is different following judgements about intelligence and physical features (Winograd, 1976) but not following judgements about intelligence and weight (Mueller et al., 1979; Winograd, 1976). According to Winograd (1976), a judgement about weight requires a more global process than a judgement about a physical feature, because weight is a global characteristic of the person. One could argue that judgements about intelligence and weight both require an holistic assessment (e.g. Woodhead et al., 1979). However, these explanations need further clarification. First, we cannot draw a conclusion from the lack of difference between recognition rate after judgements about intelligence and about weight. Second, it has been argued that judgements about intelligence lead to a high recognition rate because the processing was supposed to be holistic.

**Judgements About Gender.** In everyday life, judgements about gender could be based on simple cues (e.g. presence of a tie, make-up). Moreover, this is not possible in the laboratory when these cues are carefully discarded (e.g. Bower & Karlin, 1974; Carey et al. 1980; McKelvie, 1978, Experiment 4, 1985; Strnad & Mueller, 1977). Such a decision could then be based on several local cues (thickness of eyebrows, eye size) or on more global cues (shape of the jaw), or even on relationship between facial features as suggested by O'Toole, Millward, and Anderson (1988). If we consider that just one cue is not sufficient, such a judgement could require an holistic process (e.g. Brown & Perrett, 1993). Indirect data show that gender judgements are processed holistically. Low spatial frequencies convey information related to global characteristics of the face and are integrated by the visual system earlier than high spatial frequencies (Breitmeyer, 1984; Ericksen & Schultz, 1979). But Sergent (1986, Experiment 1) found that categorisation based on gender is performed faster than that based on profession. Moreover, gender categorisation is improved less by the presence of high spatial frequencies than profession categorisation. This indicates that gender categorisation is mainly performed through global facial characteristics (low spatial frequencies). Interpreting such a result within the componential/holistic framework, one could wonder why gender judgements always lead to lower recognition rates than personality judgements in yes/no recognition tasks (Bower & Karlin, 1974; Strnad & Mueller, 1977) and in old/new recognition tasks (Carey et al., 1980; McKelvie, 1978, 1985). Following Baddeley (1979), one could argue that this type of encoding is not appropriate to distinguish faces of the same sex at the test stage. Indeed the judgement made at the encoding
stage is a judgement distinguishing a female face from a male face. This kind of judgement involving the processing of physical features could be quite useless at the test stage in distinguishing one female face from the other female faces seen during encoding.

*Judgements About a Distinctive Feature.* Judgements about a distinctive feature, like judgements about a personality trait, lead to a high level of recognition. As mentioned earlier, these results are in agreement with the hypothesis of Winograd (1981) which contends that a judgement about a distinctive feature leads to a high level of recognition because of the great number of features processed. However, an alternative interpretation could be proposed. In only one study (e.g. Courtois & Mueller, 1979) did judgements about a distinctive feature lead to a lower recognition performance than judgements about personality. Interestingly, the list used by these authors contained no feature inducing *ipso facto* an holistic processing, i.e. chin, lips, nose, eye, and forehead. On the contrary, in the other studies the list included at least one facial feature implicating a global processing of the face: ‘‘head shape’’ (Valentine & Bruce, 1986, Experiment 3); ‘‘head shape’’, and ‘‘skin’’ (Daw & Parkin, 1981; Deffenbacher et al., 1981; Winograd, 1981, Experiments 1 and 2); ‘‘hair’’ (Parkin & Goodwin, 1983). We could argue that because of such features the nature of the processing is holistic during distinctive feature judgements as well as during judgements about personality.

**Componential/Holistic Processing and Hemispheric Specialisation**

Numerous data indicate that the left hemisphere is specialised in componential information processing while the right hemisphere is specialised in holistic information processing (e.g. Bradshaw & Nettleton, 1983; Young & Ellis, 1989). Galper and Costa (1980) examined the influence of analytic and holistic encoding on subsequent facial recognition. In the encoding phase, target faces were accompanied by *social or physical* information about each face (e.g. respectively ‘‘The woman is vivacious and outgoing’’, ‘‘This woman has curved eyebrows and thin lips’’). The test faces were presented either in the right visual field or in the left visual field. The results show important interindividual differences and an encoding by visual field interaction (see also Proudfoot, 1982, Experiment 2, for similar results). Moreover, this lateral pattern is not obtained in the standard intentional learning paradigm (Experiment 2). Therefore, faces are presumably processed preferentially by the right of left hemisphere according to the encoding type. The left hemisphere recognition performance was superior to the right one following facial feature judgements, whereas the right hemisphere was superior following judgements about personality. Such a result is quite in agreement with data showing a left
hemispheric superiority in analytic processing, and a right hemispheric superiority in holistic processing (e.g. Magaro & Moss, 1989; Van Kleeck, 1989). However the Galper and Costa study is not very convincing. First, the lateral exposure duration was too long (300ms) to prevent possible ocular movements. Galper and Costa (1980) pointed out that this duration makes the results more robust: indeed in spite of this, hemispheric differences were observed. Second, one could argue that the partly verbal encoding (through the sentences accompanying the target face) favoured the left hemisphere processing (e.g. Moscovitch, Scullion, & Christie, 1976; Servos & Peters, 1990).

Analytic Processing: A Childhood Specificity?

Some studies show that children process faces analytically, feature by feature, until 11 or 12 years old and then process faces holistically (e.g. Carey & Diamond, 1977; Flin, 1985; Flin & Dziurawiec, 1989). If children process faces in an analytic way, standard intentional learning must not lead to different recognition performance from encoding about gender. Indeed, Blaney and Winograd (1978) found a result compatible with such a hypothesis. However this hypothesis cannot be methodologically accepted because no conclusion can be inferred from an absence of difference. On the other hand, if children up to 11 or 12 years old process faces in an analytic way and are not able to adopt an holistic strategy, the recognition performance produced should not be different according to the type of encoding. But Blaney and Winograd (1978) and Carey et al. (1980) obtained the same pattern of results with children as that observed with adults; judgements about personality led to a better recognition rate than judgements about gender. Judgements about personality could allow children to adopt an holistic strategy that they do not use spontaneously. However, without a control group it is not possible to draw any conclusions.

This potential capacity, if it exists (see for example Baenniger, 1994), does not seem to vary with age. Indeed the magnitude of the improvement in recognition allowed by personality encoding compared with the recognition level obtained after feature encoding, does not seem to vary with the age of the subjects (from 7 to 14 years old; Carey et al., 1980). Thus, as children can benefit from instructions to encode faces, the analytic processing cannot be considered as a childhood specificity. Moreover, recent investigations show that the efficiency of both featural encoding and holistic encoding increase with age (Chung & Thomson, 1995; Ellis & Ellis, 1994; Pacteau & Bonthoux, 1994).

The depth of processing model and the elaboration model are difficult to test. Superior recognition performance may be assumed to arise from deeper processing, and this requires a more elaborate degree of feature encoding. Moreover, as mentioned earlier, we cannot be sure that the subject acts in accordance with instructions. However, the componential/holistic framework
offers a powerful way of examining the results. It would be possible to manipulate the stimuli in order to induce a componential or a holistic processing of the face. More precisely, we could ask subjects to make judgements about personality or about a facial feature, and to manipulate the test face (presented during the recognition test) in order to control the type of processing carried out. For instance, the test face could be presented upside down, so interfering with configurational (and holistic) processing, or it could be blurred, so interfering with componential processing. The test face could also be filtered, either containing high-pass spatial frequencies (conveying fine details) or low-pass spatial frequencies (conveying coarse information) (see Coin, 1993 for a presentation of the spatial frequency approach on face recognition).

IMPROVEMENT OF FACE RECOGNITION

Type of Processing Considered Most Effective by Subjects

Before examining the possible improvement of face recognition, one could ask subjects themselves. Irrespective of their effective way of encoding faces, which is the most effective strategy, as they see it? To look carefully at a facial feature could improve face recognition. This assertion, considered as conveying the commonsense view (Courtois & Mueller, 1979), seems to be confirmed in an informal way by Winograd (1978, 1981; see also Mueller & Wherry, 1980). Indeed, Winograd usually asked the introductory psychology students to guess which strategy is most effective; judging faces in terms of personality, judging faces in terms of a facial feature, or encoding faces without any judgements (that is to say in their own way). Invariably, the majority of the students voted for the feature processing strategy. Baddeley and Woodhead (1983) first separated subjects into two groups according to the recognition rate obtained in a previous experiment. The first group obtained a higher level of recognition than the second group. Then three recognition tests were administered, among them a facial recognition test. The results obtained were quite in agreement with the two-class separation carried out: the first group obtained significantly superior recognition performance to the second group. However the first group subjects claimed that they had analysed faces feature by feature. This result contrasts with the data presented in this review which clearly show that judgements about personality lead to the best recognition rate. Indeed, the results of a post-experimental questionnaire (Sporer, 1991) indicated that judgements about personality, just as judgements about the most distinctive feature, are considered to be more effective than judgements about physical features. Are subjects conscious of the most effective strategy for memorising faces? The agreement between which strategy is considered to be the most effective by subjects, and the strategy that is demonstrated experimentally to be the most effective,
although sometimes observed (Courtois & Mueller, 1979; Sporer, 1991), could not be considered as complete.

**Improvement of Face Recognition**

Is it possible to improve face recognition performance? In the several studies reviewed here, it does not seem that a ceiling effect occurs. Indeed, asking subjects to make personality judgements about faces improves recognition performance, compared to the recognition rate obtained after standard instructions (cf. Table 2). On the other hand, although the human ability to recognise unknown faces is quite remarkable in laboratory conditions and in everyday life, this ability does not always reach its ultimate level. For example, this level increases with repeated displaying of faces (Ellis, Shepherd, Gibling, & Shepherd, 1988) or with an increase of exposure duration (McKelvie, 1990, Experiment 1; Reynolds & Pezdek, 1992, Experiment 1). So, which type of training could permit an improvement of subjects' facial mnesic performance?

Woodhead et al. (1979) submitted subjects to a three-day training course including lectures, discussions, and exercises with Penry's (1971) Photofit. The training consisted of observation (of stance, faces, accessories from photographs), memory (oriented towards feature selection from the Photofit kit), and verbal description (also derived from the Photofit). The results obtained are rather puzzling: no improvement (in a recognition task: Experiment 1, or in a matching task: Experiment 2), but rather a decrease of performance (in a matching task: Experiment 3) was observed. A similar result was obtained by Malpass, Lavigueuer, and Weldon (1973, Experiment 1) with verbal training.

The arguments proposed to explain such lack of improvement are as follows:

1. This negative result could be ascribed to the use of the Photofit which induces a featural processing of the face, that is, of the eye, the nose, etc. (Woodhead et al., 1979);

2. The process of recognising faces is so overlearned through life that any extra training could not significantly improve it (Baddeley & Woodhead, 1983; Deffenbacher, 1988; Woodhead et al., 1979);

3. The learning methods selected are not suitable and too artificial (Malpass, 1981; Ellis, Shepherd, & Davies, 1975).

It is not surprising that a training procedure favouring the analytic processing of a face (e.g. Woodhead et al., 1979) does not permit an improvement of memorisation. Such a strategy, induced by encoding instructions, leads to a lower mnesic performance than any other type of encoding instruction (standard

---

3 A Photofit enables the composition or reconstruction of a face by superimposition of transparencies, on which the main facial features are represented.
learning, judgements about personality or about a distinctive feature). One could suspect that verbal training (e.g. Malpass et al., 1973) does not improve face memorisation either. Indeed, the quality of the verbal descriptions does not correlate with the capacity for face identification (Deffenbacher, 1988; Pigott & Brigham, 1985). Verbalisation of faces has been shown to have only a few positive effects on recognition performance (Christie & Ellis, 1981; Klatzky & Forrest, 1984, Experiment 3), and even a negative effect (Schooler & Engstler-Schooler, 1990). Subjects themselves judge verbal description of a face to be more difficult than memorisation (Laughery, Duval, & Wogalter, 1986). Moreover, to verbalise only slightly improves recognition performance (Bruyer, 1982). However, verbal descriptions of faces generally lead to a lower memory performance than reconstruction techniques using Photofit and Identikit for criminal investigations (e.g. Davies, 1983), which in turn lead to a lower memory performance than recognition tasks (see e.g. Tiberghien, 1983, for a description).

As mentioned earlier, judgements about personality traits lead to higher recognition performance than standard learning instructions. Would such a strategy within a training course permit an improvement of face recognition? Malpass (1981) submitted subjects to various training sessions. The subjects of the first group were trained to analyse facial features and their various changes through Penry’s (1971) Photofit. The second group of subjects were asked to rank faces according to personality characteristics. The third group of subjects were asked to look carefully at the general physical appearance of faces, and then to rank them according to their similarity with four other faces. Four successive recognition tasks were administered to the fourth group. No difference to a subsequent yes/no recognition test occurred whatever the previous training.

It is possible, as Malpass (1981) pointed out, that the subjects’ ability in encoding and recognising faces is so high that any training method to improve it has no effect. Moreover the methods themselves could be too crude, in view of natural strategies developed by subjects. Finally, we could argue that these courses are perhaps based on inadequate hypotheses of how people process faces. The Malpass (1981) experiment shows that training based on holistic processing of the face does not improve recognition performance. One could point out that some people could spontaneously process faces holistically. Mueller and Thompson (1986, Experiment 1; see also Mueller & Thompson, 1988) asked subjects to give as many personality traits as possible about figures, either before the study phase or after the test phase. False alarms were fewer for subjects who gave numerous personality traits than for other subjects. However, correct decisions for the former subjects varied between old and new faces. It is possible that these subjects adopt a high criterion which enhances performance on new faces and reduces performance on old faces. But we cannot claim that the recognition rate is higher for subjects who process faces holistically (as was
supposed in this experiment) than for others. This superiority, if it exists, seems to be observed only in recognition tasks. When subjects were asked to recognise faces from Identikit, the analytic strategy that consists of examining each facial feature led to a higher recognition level than the holistic strategy (e.g. Laughery et al., 1986).

On the other hand, Wells and Turtle (1988) noticed that judgements about personality traits lead to better recognition performance than do physical judgements. In contrast, judgements about physical features are in turn more efficient for verbal description (or for reconstruction with an Identikit: e.g. Wells & Hryciw, 1984) of the studied face (see Fig. 2).

The data presented here suggest that:

![Graph showing recognition and verbal description of faces as a function of encoding instructions.](image-url)

1. If subjects are asked to recognise faces among others (within a crowd or in an identification parade for criminal investigation, the most efficient encoding strategy is to judge faces according to inferred characteristics (e.g. personality trait).

2. When the test involves a description of the face (for example from an *Identikit* or a *Photofit*) the most effective encoding strategy consists of judging faces according to physical features.

**CONCLUSIONS**

It is well established that encoding instructions influence subsequent facial recognition. However the explanations proposed are controversial. The depth of processing theory offers an attractive framework to explain the results obtained. Nevertheless it has been challenged by the Winograd (1981) elaboration hypothesis. Unfortunately the available data do not enable us to say exactly whether it is the depth of processing, or the number of features encoded (elaboration hypothesis) that leads to better recognition performance than the process of one facial feature. Results showing that judgements about personality permit a better subsequent recognition rate than judgements about physical features have been obtained in various situations. As several authors have claimed, it seems that the former would be holistic in nature, and the latter analytic. Furthermore, such a hypothesis is consistent with neurophysiological data which show that the contribution of low spatial frequencies conveying holistic aspects of the stimulus is superior to the contribution of high spatial frequencies transmitting featural information for face recognition (see Coin, 1993, for a review). Unfortunately the componential/holistic approach was only considered as a framework to interpret the results obtained. Nevertheless this approach offers two main advantages in evaluating the effect of encoding instructions on face recognition. First, it is possible to manipulate the stimuli in order to control the information available to be either holistic by blurring (or by a low-pass spatial frequency filtering), or componential by presenting the face upside down (or by a high-pass spatial frequency filtering). Thus, we could obtain an equivalence between the type of encoding requested and the type of encoding performed, which is difficult within the depth of processing and the elaboration models.

Judgements made during encoding could make the retrieval of facial information stored in memory easier. However, this facilitation depends both on the type of judgement made and on retrieval conditions. For example, a judgement about a facial feature is more effective than a personality judgement in facial description or reconstruction tasks. Relative to standard intentional learning conditions, to ask subjects to make judgements about a physical facial feature disrupts the subsequent recognition rate. On the contrary, to ask subjects to make judgements about personality enhances the recognition rate. From these
results, is it possible to infer that facial recognition ability, in spite of its high degree of efficiency, could be improved? However, the training courses intended to improve this ability did not show convincing results. The failure of the training procedure is probably due to the artificial situation of face memorisation in the laboratory. One could assume, on the other hand, that several factors are involved in face memory in everyday life, depending on the situation. For example, the reliability of eyewitness identification could be improved by context reinstatement procedures (e.g. Cutler, Penrod, & Martens, 1987) although context cues could also have an important role in laboratory research (e.g. Davies, 1988; Tiberghien, 1986).

Finally, given that accuracy is usually so high, any improvement might have been masked by a ceiling effect (Baddeley, 1979). However, facial recognition ability could probably be increased by improving recognition latency. Some studies have shown that low spatial frequencies are integrated faster than high frequencies by the visual system (Breitmeyer, 1984; Ericksen & Schultz, 1979). Moreover, high frequencies of a stimulus could be reduced by blinking. Thus to blink or to screw up one’s eyes could be an easy way to favour the low spatial frequency processing (conveying holistic information and integrated faster by the visual system) allowing an holistic and fast processing of a face.

Manuscript received November 1995
Manuscript accepted 31 July 1996

REFERENCES


