Temporal integration in face perception: Evidence of configural processing of temporally separated face parts

Article *in* Journal of Experimental Psychology Human Perception & Performance · March 2007

DOI: 10.1037/0096-1523.33.1.1 · Source: PubMed

3 authors, including:

David Anaki
Bar Ilan University

Morris Moscovitch
University of Toronto

Journal of Experimental Psychology:
Human Perception and Performance
2007, Vol. 33, No. 1, 1–19

0096-1523/07/$12.00 DOI: 10.1037/0096-1523.33.1.1

Temporal Integration in Face Perception: Evidence of Configural Processing of Temporally Separated Face Parts

David Anaki
Rotman Research Institute

Jennifer Boyd
University of Toronto

Morris Moscovitch
Rotman Research Institute and University of Toronto

Temporal integration is the process by which temporally separated visual components are combined into a unified representation. Although this process has been studied in object recognition, little is known about temporal integration in face perception and recognition. In the present study, the authors investigated the characteristics and time boundaries of facial temporal integration. Whole faces of nonfamous and famous people were segmented horizontally into 3 parts and presented in sequence, with varying interval lengths between parts. Inversion and misalignment effects were found at short intervals (0–200 ms). Moreover, their magnitude was comparable to those found with whole-face presentations. These effects were eliminated, or substantially reduced, when the delay interval was 700 ms. Order of parts presentation did not influence the pattern of inversion effects obtained within each temporal delay condition. These results demonstrate that temporal integration of faces occurs in a temporary and limited visual buffer. Moreover, they indicate that only integrated faces can undergo configural processing.

Keywords: face perception, temporal integration, visual memory, configural–featural processing

The visual information produced by the retina for further analysis is fragmented, discontinuous, and nonselective, differing markedly from the orderly world that humans perceive. Perceptual organization is therefore an essential prerequisite for coherent unified visual information to be structured from the myriad of retinal piecemeal input. One process crucial for perceptual organization is temporal integration, which combines temporally separated visual components into a unified representation. Because many scenes and objects are partly occluded or too complex to be grasped in a single glance, a mechanism that retains previously acquired representations and incorporates within them new visual input, acquired across saccades, is vital. The aim of the present study was to investigate the characteristics of temporal integration of facial stimuli and to consider the ramifications of the findings for theories of face perception and visual temporal integration.

Visual Temporal Integration in Scene and Object Perception

The concept of temporal integration is linked closely to the ability of the human perceptual system to construct a temporary memory store in which incoming sensory information can be accrued. The issue of the quality and quantity of visual information that can be preserved from one view to the next has been intensively investigated in the literature, in domains other than face recognition, and is a matter of current debate (for reviews, see Henderson & Hollingworth, 2003; Hollingworth, 2004; Irwin, 1993; Rensink, 2002; Simons, Mitroff, & Franconeri, 2003; Simons & Rensink, 2005). We now summarize briefly the different views on the issue and consider their relevance to the processes involved in face perception.

Traditional accounts assumed that visual information can be accumulated across saccades and that a coherent image can be formed after adjustments are made to the changed retinal positions of the image from fixation to fixation (e.g., Breitmeyer, 1984; Feldman, 1985; Jonides, Irwin, & Yantis, 1982; McConkie & Rayner, 1976; Trehub, 1977). One illustrative example of these models is the *integrative visual buffer* model, proposed by McConkie and Rayner (1976), which presumes the existence of an iconic memory store in which the detailed contents of a fixation are temporarily held. Following a saccade, the visual information from the new fixation is aligned and superimposed with the previous one on the basis of their spatiotopic coordinates (for reviews on the mechanisms of this spatiotopic calibration, see also Bridgeman, van der Hejiden, & Velichkovsky, 1994; McConkie & Currie, 1996).

David Anaki, Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada; Jennifer Boyd, Department of Psychology, University of Toronto, Toronto, Ontario, Canada; Morris Moscovitch, Rotman Research Institute, Baycrest Centre for Geriatric Care, and Department of Psychology, University of Toronto.

We express our appreciation to Marlene Behrmann, Malcolm Binns, Romina Palermo, and Gillian Rhodes for their helpful comments and suggestions on earlier versions of the article. We thank Marilyne Ziegler, Debbie Talmi, Irina Nica, and Aliza Goodman for their help in programming and running the experiments.

Correspondence concerning this article should be addressed to David Anaki, who is now at the Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel, or Morris Moscovitch, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada. E-mail: danaki@mscc.huji.ac.il or momos@psych.utoronto.ca

More recent claims, however, have questioned these assumptions. Indeed, there is no dispute that precategorical sensory visual information can be stored for a short duration, extending between 80 and 100 ms, in a temporary storage termed *iconic memory* by Neisser (1967; Sperling, 1960). Under these conditions, temporal integration is often found (e.g., Di Lollo, 1980; Dixon & Di Lollo, 1994; Loftus & Irwin, 1998). However, visual memory—and, consequently, integration of successive scenes—fails at longer intervals and/or following saccades. These findings have led some to conclude that internal visual representations are not formed at all and visual information is acquired from the external world, which acts as its own "outside memory" (O'Regan, 1992; O'Regan & Noë, 2001). The *coherence theory* (Rensink, 2000, 2002), for example, argues that continuity is limited and achieved only for the object that is the focus of attention, with attention acting as a consolidator of its basic features. When attention is redeployed, the object disintegrates into its elementary components, and its memory dissolves.

Others (Irwin, 1993, 1996; Irwin & Andrews, 1996; Irwin & Zelinsky, 2002) have suggested that transsaccadic memory is heavily dependent on visual short-term memory (VSTM), because both phenomena have similar characteristics, such as limited capacity (3–4 items), long duration (up to 5,000 ms), and location-independent representation (Irwin, 1991). Thus, the conscious experience of a rich and stable environment across saccades is not based on the memories of previous eye fixations but, rather, on perceptual processing during current fixations. This conclusion is plausible because the durations of fixations are tenfold longer than those of saccades.

Finally, a third approach, proposed by Henderson and Hollingworth (2003; Hollingworth, 2004), argues that visual integration draws on visual long-term memory (VLTM) in addition to VSTM in maintaining memory representations across scenes or saccades. Although these representations are not sensory but abstract, they are detailed enough to allow comparison of information obtained in the previous fixation and information obtained in the current one. More important, the ability to notice changes to an object presented earlier, despite fixations on multiple intervening objects, supports the claim that visual information can be stored over long periods of time without the apparent capacity limitations that plague VSTM (Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001).

The research from which these different theories have evolved has focused almost exclusively on scene and object perception; little is known about the construction of face representations across time. As we elaborate below, the processes underlying face and object perception are not identical, and as a result, temporal integration may have unique influences and characteristics when facial stimuli are processed. Indeed, several recent studies have focused on the immunity of facial stimuli to change blindness, using the change detection paradigm, which has become a common experimental tool with which to address the issue of visual temporal integration (Barton, Deepak, & Malik, 2003; Buttle & Raymond, 2003; Davies & Hoffman, 2002; Humphreys, Hodsoll, & Campbell, 2005; Palermo & Rhodes, 2003; Ro, Russell, & Lavie, 2001). In change detection tasks, participants are asked to detect a change across views when that change is contingent on the presentation of an additional visual transient—such as a saccade, a gap, or an eyeblink—to eliminate the detection of motion incurred by the change (e.g., Rensink, 2002; Simons & Levin, 1997). Yet, the underlying endeavor, shared by the above-cited studies, was to explore the effects of attentional factors on facial perception and, consequently, on the ability to detect changes. The nature of the representation retained across presentations and the influence of temporal integration on subsequent face perception generally have been overlooked.

A notable exception is a study by Wallis & Bülthoff (2001; see also Ikeda & Uchikawa, 1978; Singer & Sheinberg, 2006), which examined the influence of temporal contiguity on the recognition of faces. In this study, observers were shown a sequence of two faces, A and B, in which the identity of the face changed (from A to B, or vice versa) as the head rotated in depth. During a later recognition test, the participants had to compare a profile and a frontal view and determine whether the images originated from the same head or from different heads. Half of the nonmatching faces were taken from faces that were paired during training, whereas the other nonmatching pairs were faces that were not paired during training. The findings revealed that participants were more likely to judge nonmatching faces, which were presented in the same training set, as belonging to the same person than they were faces that did not appear in the same training set. These results demonstrate that the visual system uses temporal contiguity to construct mental representations of faces. Because different views of a face are often seen in rapid succession, temporal correlation may serve as a powerful instrument by which to obtain a detailed image of a human face. Yet, Wallis and Bülthoff's study focused on the integration of a complete face from different viewpoints. Basic questions that await answers are whether temporal integration occurs when a single face is presented in a piecemeal fashion and what the consequences are of the success or failure of temporal integration to the processes underlying face perception.

Processes Involved in Face Perception

The role of visual temporal integration in face processing also may have important implications for theories of face perception. Most theories of face recognition distinguish between analytic or part-based processes, characteristic of object perception or inverted faces, and holistic ones that typify perception of upright faces (for reviews, see Farah, Wilson, Drain, & Tanaka, 1998; Gauthier, Behrmann, & Tarr, 1999; Gauthier & Tarr, 2002; Kanwisher & Moscovitch, 2000; Kimchi, 1992; McKone, Martini, & Nakayama, 2003; Moscovitch, Winocour, & Berhmann, 1997; Peterson & Rhodes, 2003; Tanaka & Farah, 2003). Although this holistic–analytic distinction is fundamental in the face and object recognition literature, its exact definition and operationalization have proven to be elusive. One prevalent interpretation, the *template* hypothesis, asserts that in part-based processes, the stimulus is identified on the basis of its constituent parts, whereas in holistic processing, no decomposition processes are involved—rather, the stimulus is apprehended and represented as a perceptual *gestalt*, or template, without being constructed from the representations of its basic parts (e.g., Farah, Tanaka, & Drain, 1995; Tanaka & Farah, 2003). An alternative approach emphasizes the spatial–relational information between the stimulus' parts as a crucial component in face processing (Cooper & Wojan, 2000; Diamond & Carey, 1986; Rhodes, 1988). According to this widespread view, termed the *configural* account, face perception depends on the computation of precise distances between internal face parts (e.g., interocular distance). In the present article, we adopt this view to characterize face perception processes and their sensitivity to temporal integration, deferring the implications of our findings on the other account to the General Discussion.

A predominant phenomenon that demonstrates the divergent processes involved in face and object recognition is the *inversion* effect, which is the difficulty humans have in recognizing an inverted face compared with an upright face. In contrast to the ease with which inverted objects are recognized, humans experience great difficulties in recognizing upside-down faces (Diamond & Carey, 1986; Valentine, 1988; Yin, 1969). Many accounts suggest that the encoding of configural information is disrupted when a face is inverted, and the perceiver has to resort to an analytic process that relies more on the components' information embedded in the face (but see Murray, 2004, who suggested that inverted faces are processed configurally to some extent). Objects, however, are processed in the same manner regardless of orientation and, thus, suffer to a lesser degree from their inversion (Bartlett & Searcy, 1993; Leder & Bruce, 1998, 2000; Maurer, Le Grand, & Mondloch, 2002; Rhodes, Brake, & Atkinson, 1993; Searcy & Bartlett, 1996; Tanaka & Farah, 2003; Tanaka & Sengco, 1997; Young, Hellawell, & Hay, 1987).

An important question, which largely has been ignored in the literature on face perception, is whether all of the stimulus information needs to be presented simultaneously for configural processing to occur or different components can be presented in piecemeal fashion over a short period of time and still lead to the perception of a face as a whole. That is, can faces still be perceived configurally if the presentations of different components of the face are separated by short temporal intervals? A positive answer to this question would suggest that configural processing does not depend on the co-occurrence of components but, rather, on the integration of separate components while they are held in some form of temporary storage or buffer. Farah et al. (1995) addressed this issue, to some extent, by asking participants to study faces for a subsequent memory test. The faces were presented either in a whole version or in a part version (in which each face part [eyes, nose, and mouth] was shown on a separate sheet of paper). Later, an old–new recognition test was administered, with upright or inverted whole faces presented. The results revealed that the presentation of an inverted face during the test had a detrimental effect only for faces that were studied intact, not for faces whose features were shown separately at study. Farah et al. (1995) interpreted these results as evidence for the claim that holistic processes do not involve conscious part decomposition. Therefore, only faces that were studied as a whole could produce an inversion effect, which is a marker (albeit an indirect one) of the breakdown of holistic processing and the transition to analytic, part-based processing. However, no temporal interval between the parts was manipulated systematically, and the wide spatial distribution of the face parts (presented on different sheets of paper) may have precluded integration and, consequently, configural processing. The possibility that face parts could be integrated within a shorter time window is still viable, and the question of whether integration could support configural processing still awaits an answer.

In the present study, we investigated temporal integration in face processing. To this end, we compared recognition of nonfamous (Experiment 1) and famous (Experiments 2 and 3) faces that were presented either as a whole figure or in a piecemeal fashion (corresponding roughly to the eyes, nose, and mouth sections, with varying temporal intervals between the three parts).

The fact that faces can be recognized, to some degree, by focusing on individual components in the display may obviate the need for temporal integration of components to form a holistic representation. However, if this strategy underlies performance of the current task, it will be detected by the changes in the inversion effect, which is considered to be diagnostic of configural perception for faces. Consequently, we compared performance for upright versus inverted or misaligned faces (in Experiment 1) to determine the extent of temporal integration of all the components. Thus, the indication of successful temporal integration leading to configural perception in the present study was not the recognition performance per se but, rather, the magnitude of the inversion effect. The sequential presentation of face parts, with different intervals between the presentations of each part, was intended to allow us to determine whether configural processing is possible when components are presented separately or requires the simultaneous presentation of all parts. Variation of the interval should enable us to estimate the temporal boundaries over which these integration processes can occur and to address the locus of the processes involved in this integration by relating the results to the literature on temporal integration for scenes and objects.

Experiment 1

In Experiment 1, faces of unfamiliar people were presented either wholly or in parts. In the latter condition, faces were partitioned into three parts, which were presented sequentially. After the presentation of each face, participants were required to identify the face from an array of three faces presented in an upright, inverted, or misaligned position. Because inversion may disrupt the processing of facial features themselves (e.g., Moscovitch & Moscovitch, 2000), we also looked in Experiment 1 at misalignment effects, which are known to affect the configural aspects of a face but not its facial features (Moscovitch et al., 1997; Young et al., 1987). An advantage in the perception of an upright over an inverted or misaligned face was taken to be an index of holistic or configural processing. To investigate the time frame over which this potential integration process occurs, we varied the intervals between presentations of the face parts from 0 to 700 ms.

Method

Participants. Eighty-four undergraduate students at the University of Toronto (Toronto, Ontario, Canada) participated in the experiment for course credit. All participants had normal or corrected-to-normal vision.

Materials. The critical stimuli consisted of 80 Caucasian face pictures (half male, half female) from the Max Planck Institute for Biological Cybernetics (Tuebingen, Germany [http://faces.kyb.tuebingen.mpg.de]) database (Blanz & Vetter, 1999; Troje & Bülthoff, 1996). The faces were in frontal-view position, with a neutral expression and without makeup, accessories, or facial hair. The original color pictures were converted into a 256 grayscale format (74 dpi) and extended 8.79 × 8.79 cm.

For each of the faces in the critical stimuli, a piecemeal version was generated consisting of the divided parts of the face. The faces were segmented into three parts, each including a salient facial feature (see Figure 1). The top segment included the upper part of the head and was sliced just below the eyes. The middle segment included the nose and was sliced just above the lips. The bottom segment contained the mouth and

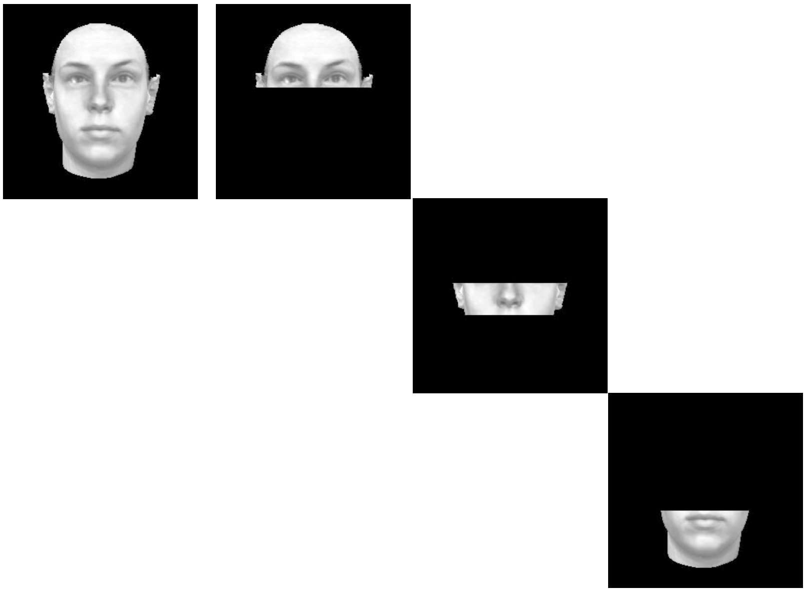

Figure 1. Example of stimuli material (whole-face and part-face conditions) used in Experiment 1. The image is from the face database of the Max Planck Institute for Biological Cybernetics (Tuebingen, Germany [http://faces.kyb.tuebingen.mpg.de]; Blanz & Vetter, 1999; Troje & Bülthoff, 1996).

was cut below the chin. In the part-face conditions, each part was presented for 17 ms, with an interval of 0, 17, 200, or 700 ms between each part (referred to hereafter as *0p*, *17p*, *200p*, and *700p*, respectively). The eyes part was always presented first, followed by the nose part, with the mouth part presented last. The parts appeared on the screen in their correct location. Thus, for example, the eyes part appeared in the same location as the eyes in the whole-face conditions.

Three whole-face conditions, in which the entire face was presented for 17, 50, or 400 ms, were added as controls (referred to hereafter as *17w*, *50w*, and *400w*, respectively). The two short-duration presentations were intended to resemble the duration of the face parts. The 400w condition was designed to serve as a control condition, in which maximal performance of inversion and misalignment effects were expected to appear.

Procedure. Twelve participants were randomly assigned to each of the seven conditions (three whole and four part conditions). Participants were tested individually, seated approximately 50 cm from a computer screen. Stimuli were displayed on an IBM color monitor controlled by E-Prime software (Psychological Software Tools, 2000) implemented on an IBM PC–compatible computer.

Each trial began with a 1,000-ms fixation mark (+) presented at the center of the screen. Following the offset of the fixation mark, an upright whole face or a combined face—composed of three parts—appeared, followed by a black-screen interval. To eliminate effects of afterimages or other types of visual persistence, we presented a mask for 500 ms in the area in which the whole or combined faces were presented. The mask was created using minute pieces of facial features taken from different faces (see Figure 2). The stimulus onset asynchrony (SOA) between the face appearance (as a whole or its first part) and the mask appearance was fixed at 650 ms for six out of seven conditions (in the 700p condition, the SOA was prolonged to 1,650 ms). As a result, the interval before mask presentation fluctuated between 200 ms (for the 200p and 700p conditions) and 633 ms (for the 17w condition).[1] The potential consequences of a variable interval before mask presentation are addressed in Experiment 2B.

Following the mask, a display of three faces was presented. The display consisted of the previously presented face as target and two other faces as distractors, selected from among the critical stimuli used in the experiment. The faces in the display were presented randomly in an upright, inverted, or misaligned orientation. The misaligned faces were constructed by dividing each face into two parts by slicing it under the eyes. The nose in the lower segment was positioned under the left ear of the upper segment.

Participants were asked to select the face in the display that was identical to the target face. They were required to respond by pressing one of three keys corresponding to left, middle, and right faces in the display. The experiment consisted of a total of 240 trials, divided equally across the three orientation conditions (upright, inverted, misaligned). In each of these orientation conditions, every critical stimulus was presented once as a target and twice as a distractor in other trials. The target appeared equally often in the three possible locations in the display.

A set of 15 practice trials was administered prior to commencement of the experiment. These trials were constructed with the same constraints as the experimental trials outlined above, and their results were not analyzed. The faces used in the practice trials were not used in the experiment but were taken from the same source. Feedback was given to participants during practice but not during the experiment.

Results

The mean accuracy of the participants as a function of condition is presented in Table 1. Below, we first present the effects of the whole–part presentation on the recognition of upright and inverted faces, followed by the effects of these conditions on misaligned faces.

Inversion effects. As shown in Table 1, upright faces were recognized better than inverted faces. The advantage of upright over inverted faces, however, was mostly noted in the whole conditions and in the short- to intermediate-interval part conditions, diminishing appreciably when the interval exceeded 200 ms (see Figure 3). These observations were confirmed by a two-way analysis of variance (ANOVA), performed with one between-

---

[1] The intervals in the other conditions were 250, 600, 600, and 570 ms for the 400w, 50w, 0p, and 17p conditions, respectively.

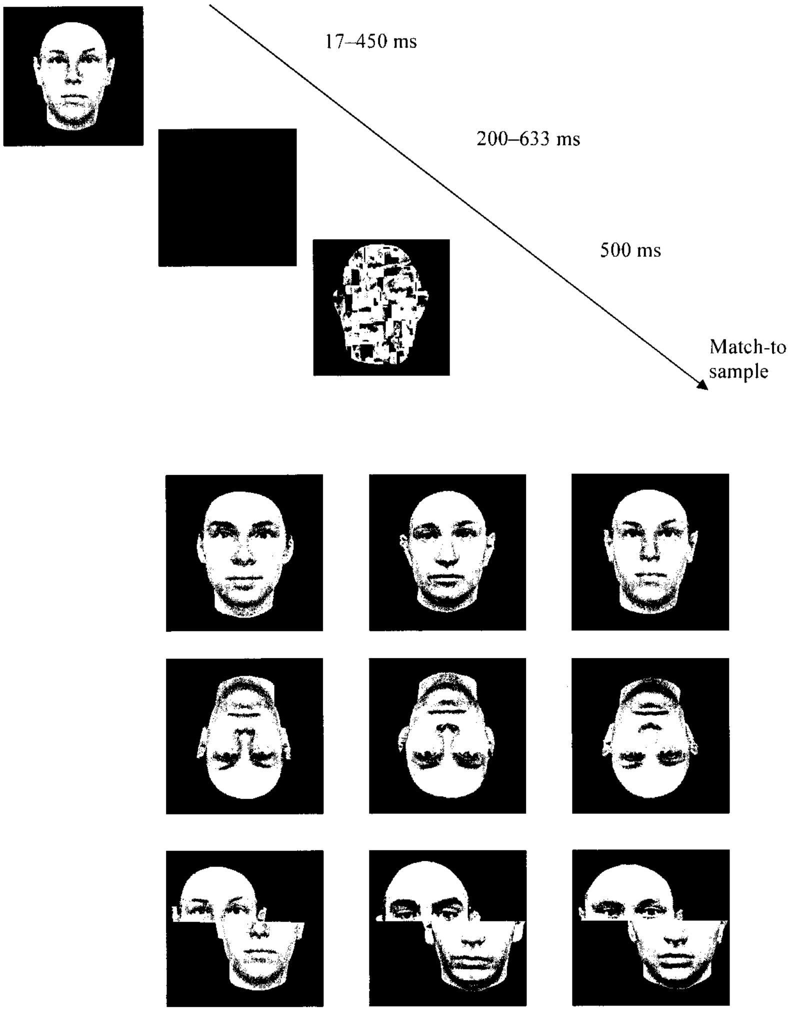

Figure 2. Sequence of events in a typical trial in Experiment 1. A target face was presented centrally either as a whole or segmented into parts, with varying interval lengths between the parts, followed by a mask. After the presentation of the face, the participant was required to perform a match-to-sample task from an array of three upright, inverted, or misaligned faces. The image is from the face database of the Max Planck Institute for Biological Cybernetics (Tuebingen, Germany [http://faces.kyb.tuebingen.mpg.de]; Blanz & Vetter, 1999; Troje & Bülthoff, 1996).

subjects factor (presentation type: 400w, 50w, 17w, 0p, 17p, 200p, 700p) and one within-subject factor (orientation: upright, inverted), on participants' mean accuracy. This analysis yielded a significant main effect of orientation, $F(1, 77) = 169.90$, $MSE = 0.003$, $p < .0001$, resulting from greater accuracy to upright (.75) than to inverted faces (.65). A main effect of presentation type was also found, $F(6, 77) = 8.10$, $MSE = 0.017$, $p < .0001$. Finally, and most important, the Presentation Type × Orientation interaction was also significant, $F(6, 77) = 4.00$, $MSE = 0.003$, $p < .005$.

We performed several analyses to examine this interaction. First, simple main effects analyses, conducted for each of the presentation type conditions between upright and inverted face presentation, revealed significant inversion effects in all conditions (all ps $< .0001$) except 700p, $F(1, 77) < 1$, $MSE = 0.003$, $p > .36$. We then conducted a Bonferroni multiple comparisons test to examine more specifically whether the four part conditions differed among themselves.[2] The test revealed that the inversion effect in the 700p condition was significantly different from that in all other part conditions (all ps $< .01$). The inversion effects of the latter conditions, however, did not differ among themselves or from those in the whole-face presentation conditions.

Misalignment effects. Accuracy was higher for upright faces than for misaligned faces, but as with inverted faces, this was seen only for whole-face and short-interval part-face presentations (see

---

[2] The Bonferroni test performed was a modified version that is based on the assumption that several comparisons ($n = df$) could be made without correcting the alpha level. Because we were not interested in making all the 21 comparisons performed in the regular Bonferroni test, we adopted a less stringent correction in which $\alpha = .01$ (for the equation used to compute the alpha level and further discussion, see Keppel, 1982, pp. 147–148).

Table 1

Proportions of Correct Responses (With Standard Deviations in Parentheses) for Upright, Inverted, and Misaligned Faces as a Function of Presentation Type in Experiment 1

	Whole-face presentation			Part-face presentation			
Condition/effect	400 ms	50 ms	17 ms	0 ms	17 ms	200 ms	700 ms
Upright	.91 (.05)	.79 (.12)	.79 (.09)	.75 (.07)	.71 (.10)	.69 (.10)	.63 (.15)
Inverted	.78 (.06)	.64 (.11)	.66 (.13)	.64 (.05)	.62 (.11)	.58 (.07)	.61 (.12)
Misaligned	.88 (.06)	.74 (.11)	.73 (.11)	.69 (.05)	.68 (.08)	.67 (.08)	.65 (.13)
Inversion effect	.13 (.05)	.15 (.08)	.13 (.10)	.11 (.05)	.09 (.04)	.10 (.08)	.02 (.08)
Misalignment effect	.03 (.04)	.05 (.06)	.06 (.08)	.06 (.07)	.03 (.07)	.01 (.08)	−.02 (.06)

Note. The inversion effect was computed by subtracting the inverted-faces performance from the upright-faces performance for each participant. The same procedure was performed for the misalignment effect (upright − misaligned). The 400-ms, 50-ms, and 17-ms columns in the whole-face presentation conditions denote the duration of the face's presentation. The 0-ms, 17-ms, 200-ms, and 700-ms columns in the part-face presentation conditions refer to the different time intervals between face parts.

Table 1; see also Figure 3). These observations were corroborated by an ANOVA conducted on the upright and misaligned target faces as a function of presentation type (400w, 50w, 17w, 0p, 17p, 200p, 700p). A misalignment effect was found, with greater accuracy for upright (.75) than for misaligned faces (.72), $F(1, 77) = 16.86$, $MSE = 0.002$, $p < .0001$. A main effect of presentation type was also found, $F(6, 77) = 10.02$, $MSE = 0.016$, $p < .0001$. The Presentation Type × Orientation interaction was also significant, $F(6, 77) = 2.25$, $MSE = 0.002$, $p < .05$.

Simple main effects analyses indicated greater accuracy for upright than for misaligned faces in all conditions (all ps $< .05$) except 200p and 700p, in which accuracy failed to reach significance. A Bonferroni test for inversion effects revealed that the 700p condition was different from all other conditions (all ps $<$ .01) except 17p and 200p. The latter two conditions, however, did not differ from the 0p or the three whole conditions.

Discussion

The results of Experiment 1 showed that if the intervals between the face parts were short enough, inversion effects did not differ between conditions in which a face was presented as a whole (for 400, 50, or 17 ms) and conditions in which face parts were presented with varying temporal intervals between them. A marked reduction in the inversion effect was observed only when the interval between the face parts was 700 ms. A sizeable inversion effect was obtained in the 200p condition—comparable to that obtained in the whole conditions. The misalignment effect was much smaller than the inversion effect, even for the whole conditions (.04 vs. .14, respectively, when the three whole conditions were collapsed together), making it difficult to detect variations in the misalignment effect in the part conditions. The smaller misalignment effect was probably a result of the strategies of some participants, who may have settled for encoding only specific features of a face's parts without attempting to integrate them into a whole face. Nonetheless, misalignment results showed a trend similar to that found for inversion. There were significant differences between the 700p and the whole and 0p conditions, with no differences among any of the other part and whole conditions. However, the difference in the misalignment effect between the 400w and 700p conditions was less pronounced, and the 700p condition also did not differ significantly from the 17p and 200p conditions.

Overall, the inversion and misalignment effects obtained in the various part conditions indicate that participants can store different face parts in a short-term visual buffer and integrate them into a single, unified face. The storage duration of this visual buffer seems to be at least 450 ms (the time between the onset of the first face part and the offset of the third face part in the 200p condition), but it does not exceed 1,450 ms (the time between the onset of the first face part and the offset of the third face part in the 700p condition). The most likely source of the memory system in which this integration occurs is considered in the General Discussion.

The results of Experiment 1 indicate that configural processing can be achieved even when all components are not presented simultaneously, as long as the separate components can be held in a short-term store while the integration process occurs. That inversion and misalignment effects in all part conditions but 700p were equivalent to those in the whole conditions suggests that the representation created from the segregated parts resembled the representation in the whole-face conditions. In other words, temporal integration of face parts in the visual short-term buffer results in a visual representation that allows configural processes to proceed. This representation is not just a piecemeal registration of the constituent parts of the face.

Further evidence for the existence of such an integrative system is provided by recent findings focusing on the temporal and spatial dynamics of face perception. Of great relevance is a recent study by Singer and Sheinberg (2006) focusing on the temporal dynamics of the *composite face* effect (Young et al., 1987). This effect stems from the difficulty in identifying a familiar half face when it is combined, in a complementary fashion, with another half face compared with a condition in which the two halves are inverted or misaligned. Singer and Sheinberg found that this effect persisted even when the two face halves were separated by an 80-ms temporal interval of visual noise. These results are compatible with the present findings in demonstrating that configural effects can evolve even when face parts are not presented simultaneously. The mechanism that allows these face-specific effects to emerge is the integration process, without which configural processing would not be possible. As demonstrated in Experiment 1, the inversion and the misalignment effect are eliminated at the long-interval condition.

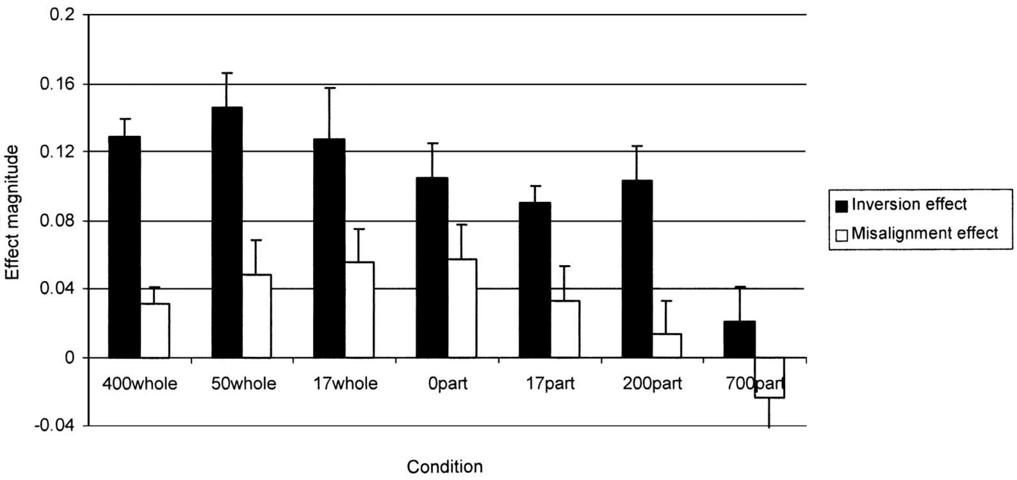

Figure 3. Magnitude of the inversion and misalignment effects in Experiment 1 as a function of presentation type. Error bars represent standard errors.

Similar findings have been reported in the spatial domain. Yovel, Paller, and Levy (2005) have shown in a match-to-sample task that accuracy for consistent hemifaces (symmetrical right and left halves that are mirror images along the midline) is greater than the summated accuracy for left and right hemifaces presented in isolation. This superiority was obtained for upright but not for inverted hemifaces, demonstrating that visual integration is a prerequisite only for configural processing. In the same vein, Gauthier, Tanaka, and Brown (2006) have reported that configural processing is obtained even when face parts are separated in space. In their study, participants were presented sequentially with two aligned or misaligned composite faces, consisting of top and bottom parts taken from different faces. One of the parts in the second (test) face was cued. Participants had to judge whether the cued part in the test face was identical to the corresponding part in the study face. The effect of the noncued part on the same–different judgment was investigated. Surprisingly, despite the fact that the face was not presented as a whole during encoding (i.e., misaligned), the authors found a misalignment effect in the second face (Young et al., 1987)—namely, the noncued part in the test face interfered more with the task when it was aligned rather than misaligned with the other part.

In conclusion, recent findings strongly suggest that temporal and spatial integration are prevalent processes in face perception. Moreover, they indicate that configural effects can arise only on integrated facial representations, making integration a necessary condition for configural processing of faces.

Experiment 2

The results of Experiment 1 indicate that integration of temporally separated face parts allows configural processes to evolve, as long as the interval between the instances of facial information to be integrated does not exceed a critical limit of approximately 400–500 ms. Our finding of inversion effects (and, to a lesser degree, misalignment effects) at the shorter intervals in the part conditions, which are comparable to the whole conditions, suggests that temporal integration occurs even when a face is presented as a whole. The purpose of Experiment 2 was to determine whether a similar pattern of results indicative of temporal integration would be found for famous faces. Unlike perception of unfamiliar faces, which may rely on effortful integration of unfamiliar elements, perception and recognition of familiar faces may be driven by top-down processes based on well-established internal representations. As a result, familiar faces may be mapped onto those representations directly (template matching), with little, if any, recourse to temporal integration. In such circumstances, configural processes may be greater for faces presented as wholes rather than as parts, requiring them to be temporally integrated over short time intervals. Indeed, part presentations could also trigger the activation of the face's representation in long-term memory, yet its activation may be weaker and ambiguous because of the partial information supplied by the part. In addition, if the identification of the face in the part condition is independent of integration and results from matching of the visual information to an existing representation, no difference would be expected between the different part conditions, unlike the findings obtained in Experiment 1. If, however, temporal integration occurs early in perception, then effects similar to those seen in Experiment 1 should be observed in identification of familiar faces.

A second, related purpose of Experiment 2 was to use identification, rather than perceptual matching, as a measure of performance. Because the sine qua non of face perception is identification, it is important to establish that any observed temporal integration effects are also applicable to identification.

In Experiment 2, we presented faces of famous people either as wholes or sequentially by parts. Both types of faces were presented either upright or inverted. A misaligned condition was not used because it would have resembled the sequential part condition and, thus, been too confusing to present at study (it was presented only at test in Experiment 1). Participants were asked to identify the faces by name or by some individuating trait (e.g., U.S. president, star actor in the *Rain Man*).

Another issue that we wanted to address concerns the time interval between the presentation of the face and the appearance of the mask. In Experiment 1, we used a varying interval, such that the entire SOA between the initial appearance of the face (or its first part) to the onset of the mask was 650 ms (only in the 700p condition was this SOA longer). This manipulation created different intervals prior to the mask appearance, which varied between 200 ms (for the 200p and 700p conditions) and 633 ms (for the 17w condition). The potential consequence of a variable interval is the differential influence of the backward pattern mask on the processing of facial stimuli appearing in close temporal contiguity. The likelihood that masking differences could account for our findings in Experiment 1 is low because of the transient effects of the mask at intervals longer than 100 ms (e.g., Enns & Di Lollo, 2000). However, to rule out this alternative account, we set a fixed temporal delay of 200 ms in the four critical conditions (50w, 17w, 0p, 17p), which were characterized by a sizeable delay in Experiment 1. We tested whether this shortened interval would influence the inversion effects. The results of this manipulation are reported in Experiment 2B; in Experiment 2A, the original variable interval procedure was used.

Experiment 2A

Method

Participants. Eighty-four undergraduate students at the University of Toronto participated in the experiment for course credit. All participants had normal or corrected-to-normal vision.

Materials. The critical stimuli consisted of 180 pictures of famous people from different fields (e.g., actresses, politicians, athletes), which were selected from an original pool of 400 pictures on the basis of familiarity ratings collected in a pilot study. The faces were frontal-view images and were downloaded from various Internet sites. We attempted to select pictures that did not contain any external cues of the identity of the figure (e.g., throne, national flag in the background). All pictures were converted into to a 256 grayscale format (74 dpi) and rescaled to a size of 5.6 × 7.2 cm.

The selection of the 180 target stimuli proceeded as follows: The 400 pictures were presented to 8 undergraduate students (each participant was shown only 200 pictures), who were asked to name the person or supply any identifying information (e.g., occupation, biographical details). Pictures were first presented in an inverted orientation and, later (if recognition failed), in an upright position. Participants' responses were rated in the following manner: 1 = the inverted face was recognized, 2 = the upright face was identified, 3 = the upright face was familiar, and 4 = the face was not recognized at all. The ratings were averaged, and a cutoff point of 2.75 was chosen, resulting in 180 pictures whose average rating was 2.75 or less.

As in Experiment 1, a divided version was generated for each face, consisting of three parts. The parts corresponded to the top, middle, and bottom sections of the face, and each included a salient facial feature (eyes, nose, and mouth, respectively). An example of a whole and split-up face of former U.S. President Bill Clinton is presented in Figure 4.

Faces were presented across participants either in a whole-face fashion or divided into parts, with the three face parts presented sequentially. In the whole-face mode, three conditions were manipulated in which the face was presented for 400, 50, or 17 ms. In the part-face mode, each face part was presented for 17 ms, with four different intervals between the parts (0, 17, 200, or 700 ms). Within each of the seven conditions, 90 upright faces and 90 inverted faces were presented.

Procedure. Twelve participants were randomly assigned to each of the seven conditions (three whole-face conditions, four part-face conditions). Participants were tested in groups, with each group containing between 4–7 participants. Each participant was seated in front of a computer and

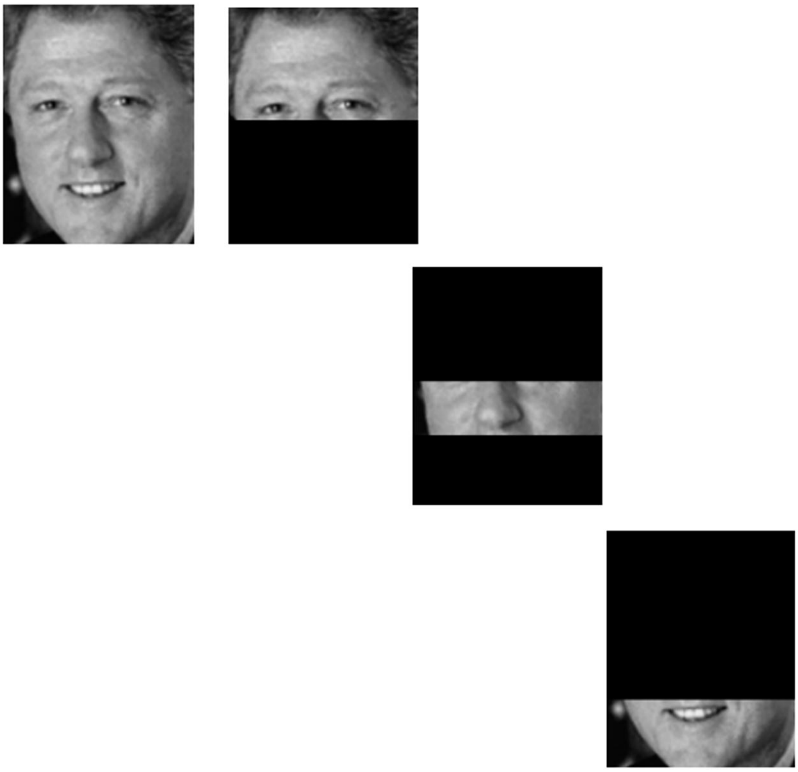

Figure 4. Example of stimuli material (whole-face and part-face modes) used in Experiments 2A and 2B. The photograph of Bill Clinton is used with the permission of Corbis Images.

individually performed the specific condition. The apparatus was identical to that of Experiment 1.

The procedure of Experiment 2 differed slightly from that of Experiment 1: Each trial began with a 1,000-ms fixation mark (+) presented at the center of the screen. Following the offset of the fixation mark, an upright or inverted famous face appeared, followed by a black screen interval ranging between 200 and 633 ms. Then, a mask appeared for 500 ms, occupying the area in which the whole or combined faces were presented. Participants were asked to observe the face and press the *1* key if the face was familiar to them or the *2* key if it was not. The keypress initiated the appearance of a unique number at the lower left side of the screen. The number was designated by the experimenter to the specific face and ranged between 1 and 180. Participants then wrote in an answer sheet, provided to them beforehand, the name of the figure or any identifying information they could recollect about the person. To advance to the next trial, participants pressed the *z* key. The sequence of the faces' presentation and the orientation of each face were determined randomly for each participant.

Following the presentation of the faces in the first part of the experiment, all faces were presented again for unlimited time in an upright condition. As in the previous part, participants were instructed to press the *1* key if the face was familiar and the *2* key if it was not. If the face was recognized as familiar, the participant wrote, in the allocated space on his or her response sheet, the name of the figure or any identifying information. The participants were asked to refrain from modifying their responses from the first part of the experiment.

Four practice trials were presented prior to the experiment itself. The four faces were first presented in an upright or an inverted orientation, in one of the seven conditions described above. Then, the faces were presented for unlimited time in an upright position. The results of these trials were not included in the analysis.

Results

The performance of each participant varied as a function of both the condition in which he or she was tested *and* his or her individual familiarity with the faces, regardless of presentation mode. For this reason, we introduced the second part of the experiment, in which faces were presented in an upright position for unlimited time and participants' knowledge was probed. Using these scores, we computed the accuracy of each participant not relative to the total number of faces presented in a specific condition but relative to the number of faces the participant recognized in the second part of the experiment. Thus, for example, if a participant identified 60 (out of 90) upright faces in the first part of the experiment and 75 faces in the second part, his or her accuracy was computed as .80, not .67. ANOVAs performed on the number of faces recognized in the second part of the experiment revealed that the level of recognition was similar across conditions both for faces that were presented initially in an upright condition ($F < 1$) and for faces that appeared inverted, $F(6, 77) = 1.19$, $MSE = 151.55$, $p > .32$ (see Figure 5). Two experimenters scored the response sheets. A response was considered accurate if the participant supplied the name of the famous figure (e.g., *Clint Eastwood*) or gave specific identifying information attesting to his or her knowledge of the person (e.g., *the Good in the movie* The Good, the Bad and the Ugly).

The mean accuracy of participants as a function of condition is presented in the top half of Table 2. Upright faces were recognized better than inverted faces, the difference between the two types being equivalent across whole and part conditions until the interval exceeded 200 ms, at which point the inversion effect was diminished. These impressions were confirmed in a two-way ANOVA, with one between-subjects factor (presentation type: 400w, 50w, 17w, 0p, 17p, 200p, 700p) and one within-participant factor (orientation: upright, inverted), performed on the participants' mean accuracy. This analysis yielded a significant main effect of orientation, $F(1, 77) = 1{,}035.89$, $MSE = 0.008$, $p < .0001$, resulting from greater accuracy for upright (.66) than for inverted faces (.22). A main effect of presentation type was also found, $F(6, 77) = 21.60$, $MSE = 0.023$, $p < .0001$. Finally, and most important, the Presentation Type × Orientation interaction was also significant, $F(6, 77) = 5.81$, $MSE = 0.008$, $p < .0001$.

The inversion effect in all the presentation type conditions was significant (all *p*s < .01; see Figure 6). A Bonferroni test revealed that the inversion effect in the 700p condition differed from that in all other conditions, whereas the inversion effects in the remaining conditions were comparable (with the exception of a significant difference between the 50w and 200p conditions [$p < .005$]).

The decrease of the inversion effect in the 700p condition could not be attributed to floor effects. Inspection of the accuracy data (see Table 2) shows that the reduction of the magnitude of the

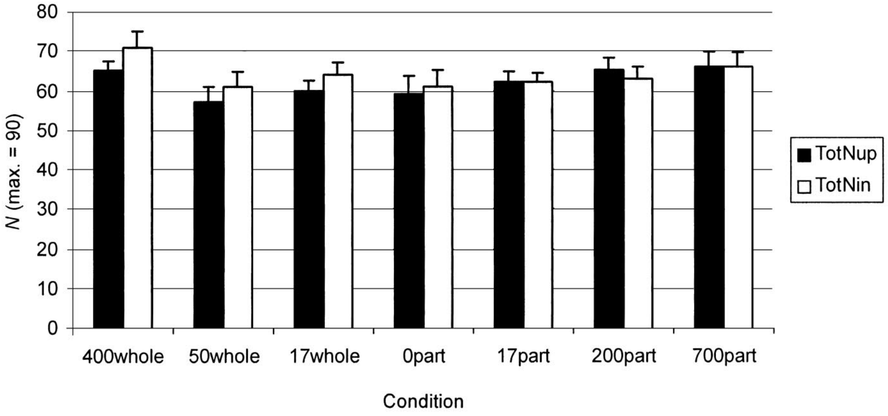

Figure 5. Total number of faces recognized in the unlimited-time upright viewing part of Experiment 2A as a function of orientation of presentation (upright or inverted) in the masked part. Error bars represent standard errors. TotNup = total number upright; TotNin = total number inverted.

Table 2
Proportions of Correct Responses (With Standard Deviations in Parentheses) for Upright and Inverted Faces as a Function of Presentation Type in Experiment 2

	Whole-face presentation			Part-face presentation			
Condition/effect	400 ms	50 ms	17 ms	0 ms	17 ms	200 ms	700 ms
				Experiment 2A			
Upright	.93 (.06)	.79 (.12)	.68 (.13)	.65 (.17)	.64 (.18)	.59 (.09)	.35 (.12)
Inverted	.47 (.20)	.24 (.09)	.19 (.05)	.18 (.10)	.18 (.16)	.16 (.08)	.09 (.09)
Inversion effect	.47 (.19)	.55 (.11)	.49 (.14)	.47 (.12)	.46 (.14)	.43 (.09)	.26 (.07)
				Experiment 2B			
Upright		.71 (.11)	.68 (.15)	.63 (.13)	.67 (.10)		
Inverted		.25 (.13)	.16 (.15)	.17 (.10)	.17 (.07)		
Inversion effect		.47 (.08)	.51 (.12)	.45 (.10)	.50 (.12)		

Note. The inversion effect was computed by subtracting the inverted-faces performance from the upright-faces performance for each participant. The 400-ms, 50-ms, and 17-ms columns in the whole-face presentation conditions denote the duration of the face's presentation. The 0-ms, 17-ms, 200-ms, and 700-ms columns in the part-face presentation conditions refer to the different time intervals between face parts.

inversion effect resulted mainly from the decrease in accuracy for the upright faces, whereas performance for the inverted faces did not appear to change. Post hoc tests, performed separately for upright and inverted faces, revealed that in upright faces, comparable accuracy was obtained for the 17w through 200p conditions, and they all differed from the 700p condition. For inverted faces, however, no significant difference in performance was found between the 700p condition and the other conditions (except 400w and 50w).

Discussion

The results of Experiment 2A extend the findings of Experiment 1 with a set of famous faces and a task that required participants to identify the face rather than choose it from among distractors. As in Experiment 1, temporal integration was observed when faces were presented in a piecemeal fashion, as indicated by the resemblance of the inversion effect in the whole and part presentation conditions up to 200p. A noticeable decrease in the inversion effect was observed only in the 700p condition. These results suggest that full temporal integration of familiar faces is achieved over the same interval as that for unfamiliar faces, indicating that configural processing of famous faces can be obtained even when the faces are presented in piecemeal fashion and identification, rather than matching, is required. Moreover, the equivalent inversion effects for the whole and part conditions obtained for famous faces support the notion that integration processes occur at early stages of perception and, as such, influence both unfamiliar and familiar face perception.

The reduced inversion effect obtained in the 700p condition suggests that the processes mediating it differed from those in the other conditions; nevertheless, the inversion effect was not eliminated entirely even at the long interval. This reduced but still

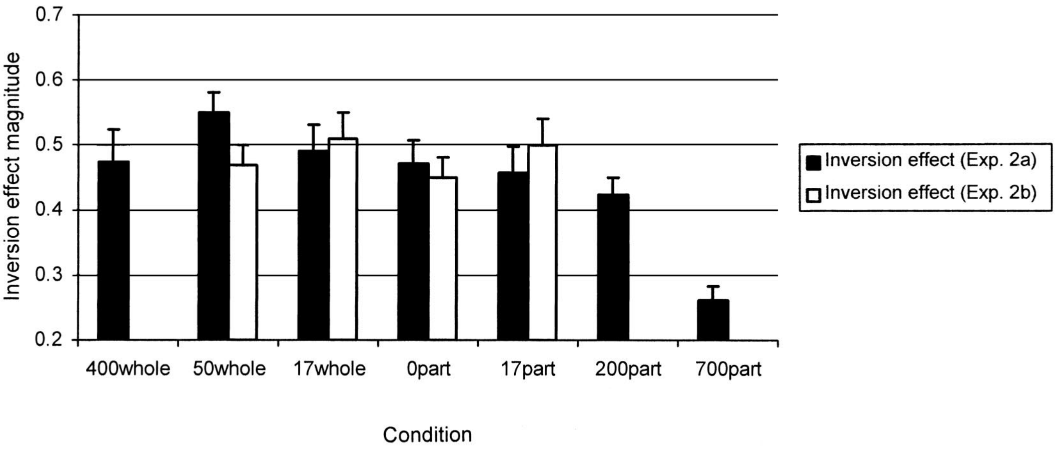

Figure 6. Magnitude of the inversion effects in Experiments 2A and 2B as a function of presentation type. Error bars represent standard errors.

significant effect may have stemmed from the inversion of featural or configural information within the part segment itself rather than from loss of integration across parts. Studies have shown that faces can be identified solely on the basis of a segment containing local relational information and that performance is degraded when face segments are inverted (Leder & Bruce, 2000). Memory representations of famous faces may be very detailed and contain rich featural and configural information on subparts of the face, particularly the region containing the eyes (Fraser, Craig, & Parker, 1990; Haig, 1986; Leder, Candrian, Hubber, & Bruce, 2001). This information, although limited, may have supported the inversion effect for famous faces in the 700p condition of Experiment 2A.

To test this possibility, we presented the eyes segment in isolation in the 700p condition. Obtaining an inversion effect in this condition would corroborate the claim that the inversion effect obtained in the 700p condition in Experiment 2A did not stem from loss of temporal integration across parts, as in the other conditions, but from inversion of the part segment itself. Eight participants performed the famous faces recognition task in a condition that was similar to the 700p condition. Only the eyes part was presented, whereas the nose and mouth parts were replaced by a black screen that appeared for the same amount of time. When the eyes section was presented in isolation, an inversion effect of .18 was obtained (upright eyes recognition accuracy was .26, compared with .08 when eyes were inverted). This inversion effect was significantly different from zero, $t(7) = 9.87$, $p < .0001$. This finding supports our conjecture that a significant portion of the inversion effect at 700p resulted from the identification of individual parts of the face and/or the local relations between them (Moscovitch & Moscovitch, 2000).

Although this loss of segment identification may account for most of the inversion effect at 700p, it does not account for all of it. Further analysis revealed that the .26 inversion effect in the 700p condition was significantly larger than the .18 inversion effect obtained with eyes only, $t(18) = 2.74$, $p < .01$. The two conditions did not differ in the total number of faces recognized, indicating a similar level of knowledge between the two groups for upright and inverted famous figures (both ts < 1). The remaining portion of the inversion effect, which was very small (.08), likely resulted from the separate and possibly independent contribution of the other parts, such as the nose and mouth (see Veuilleumier, Mohr, Valenza, Wetzel, & Landis, 2003).

The finding that significant inversion effects are obtained for face parts, such as eyes, supports the claim, advanced above, that the inversion effect obtained in the long-interval condition for famous people (700p) stemmed from configural processes occurring within the parts. The finding is at odds with the alternative hypothesis that the inversion effect over this long interval also reflects temporal integration. If the same mechanism is responsible for the inversion effects in all of the part conditions, it would be hard to account for the disproportional reduction of the inversion effect of about 40% from the 200p to 700p conditions. We tested the process discontinuity between the 200p and 700p conditions further by modeling the relationship between inversion effect size and interval length. We fitted a linear equation to the observed data in the 0p, 17p, and 200p conditions. On the basis of the regression line obtained, we extrapolated the predicted inversion effect in the 700p condition (.32) and tested whether the obtained value differed from the one predicted. The significant effect, $t(11) = 2.82$, $p < .02$, indicates that there is a process discontinuity between the two conditions, and it supports our interpretation that the factor underlying the inversion effects at 700p is different from those occurring at shorter intervals.[3]

Experiment 2B

In Experiment 2B, we investigated whether the variable interval between the face and the mask may have influenced the results obtained in Experiments 1 and 2A. To this end, we used a fixed interval of 200 ms in the four conditions that had longer delays in the previous experiments (50w, 17w, 0p, and 17p).

Method

Participants. Thirty-two undergraduate students at the University of Toronto participated in the experiment for a course credit. All participants had normal or corrected-to-normal vision.

Materials. The critical stimuli were identical to those in Experiment 2A.

Procedure. Eight participants were allocated for each of the four conditions (50w, 17w, 0p, and 17p). The procedure was identical to that of Experiment 2A with a single exception: Following the presentation of the whole face (in the 50w and 17w conditions) or the mouth part (in the 0p and 17p conditions), a black screen appeared for 200 ms, replaced later by the mask. Thus, in the four conditions of the present experiment, a fixed interval was inserted between the face and the mask, whose length was identical to that in the 200p and 700p conditions in Experiment 2A (and similar to the interval in the 400w condition, which was 250 ms).

Results and Discussion

As in Experiment 2A, the accuracy performance of each participant was computed as the proportion of the sum of correct responses in the first part of the experiment relative to the number of correct responses in the second part of the experiment, in which the faces were presented for unlimited time in an upright position. Preliminary analyses revealed that recognition levels in this latter part did not differ across the critical condition for either upright or inverted faces (both Fs ≤ 1). The mean accuracy of participants as a function of condition is presented at the bottom half of Table 2.

Our primary incentive for conducting Experiment 2B was to investigate whether the different interval before mask appearance across conditions could account for the results obtained in Experiment 2A. To this end, we conducted a two-way ANOVA, with one between-subjects factor (presentation type: 50w, 17w, 0p, 17p) and one within-subject factor (orientation: upright, inverted), on participants' mean accuracy. Only the main effect of orientation was significant, $F(1, 28) = 668.77$, $MSE = 0.006$, $p < .0001$, resulting from greater accuracy to upright (.67) than to inverted

---

[3] An alternative interpretation of the significant inversion effect obtained in the long-interval condition in Experiment 2, as compared with its absence in Experiment 1, is based not on the famous–nonfamous distinction but, rather, on the different tasks used in the two experiments. In Experiment 1, the temporal integration manipulation occurred at encoding, whereas in Experiment 2, the temporal integration manipulation occurred at the retrieval stage. Evidence from Gauthier et al.'s (2006) study suggests that configural manipulations at retrieval are more influential than those at encoding, thus leading to larger effects. We thank Isabel Gauthier for this suggestion.

faces (.19). The main effect of presentation type and the Orientation × Presentation Type interaction were not significant, the latter null effect indicating a similar magnitude of the inversion effect across all conditions. Thus, equating the delay before the mask to 200 ms did not affect the basic results seen in Experiment 2A—namely, a comparable inversion effect for the whole and part face presentation (see Figure 6).

We corroborated our conclusion with a three-way ANOVA with experiment (2A, 2B) included as a between-subjects factor along with presentation type (50w, 17w, 0p, 17p) and orientation (upright, inverted). Aside from the main effect of orientation, no effects reached significance (all Fs ≤ 1). Finally, we conducted an analysis of the four conditions from Experiment 2B and the three conditions from Experiment 2A that were not examined in Experiment 2B (200p, 700p, and 400w). This comparison is important because all of these conditions are characterized by an interval of similar length. The results replicated the ones found earlier showing that the only difference in the inversion effect was between the 700p condition and all the other conditions, with the other conditions not differing among themselves.

The results of Experiment 2B illustrate that the temporal interval length before the mask does not play a critical role in accounting for our results in Experiments 1 and 2A, because backward masking effects in face processing already dissipate when a 200-ms interval separates the face and mask. This conclusion has also been arrived at by several other researchers. Rolls and Tovee (1994) showed that when human observers are required to identify from among six faces a face that was presented and backward masked, accuracy reached a level of 97% with a brief interval of 40 ms. A lateralization study (Heider & Groner, 1997) yielded similar results, showing that a steady state of performance was obtained by a 75-ms delay (although, admittedly, the maximum delay explored in that study was 135 ms). Moscovitch and Radzins (1987), who investigated the effects of mask type, target duration, and target–mask interval, also found transient effects of face identification that disappeared early. Thus, effects of masking with face parts at 200 ms may be negligible (see also Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005).

Experiment 3

In the previous two experiments, the presentation order of the three face parts was fixed. The eyes appeared first, followed by the nose, with the mouth appearing last. It remains to be determined, however, whether temporal integration can occur even when the parts do not appear in a consecutive serial manner. On the one hand, perceptual organization is guided by grouping laws. One of them is the proximity principle, according to which elements tend to be grouped together if they are close to one another (Kubovy, Holocombe, & Wagemans, 1998; Kubovy & Pomerantz, 1981; Wertheimer, 1925). Temporal integration may, therefore, be facilitated when adjacent parts are presented sequentially. On the other hand, perception of a face may not necessarily entail a strict scanning pattern in which adjacent areas are encoded sequentially. An observer may choose to concentrate on facial features selected at random or selected on the basis of their distinctiveness. In real life, different parts of a face may be occluded, forcing the perceiver to integrate noncontiguous regions. Eye movement studies also have shown that although some consistency can be seen in the saccades performed during face perception, variation in scanning strategy occurs as well (Walker-Smith, Gale, & Findlay, 1977). Thus, temporal integration may be achieved even when the face parts are not presented in an ordered sequence.

Experiment 3 may also shed light on the type of memory system involved in temporal integration of faces. Of the two likely contenders, iconic memory and VSTM, the former is less influenced by configural grouping than the latter (e.g., Hollingworth, Hyun, & Zhang, 2005; Irwin, 1991). If presentation-order effects are found (especially in the 200p condition), VSTM is favored as the locus of the temporal integration. In contrast, if no order effects are found, iconic memory is favored as the probable locus.

In Experiment 3, the presentation order of the three face parts was varied. Thus, in addition to the original order of eyes, nose, mouth (ENM) used in the previous two experiments, five other possible combinations of these parts were tested as well (EMN, NEM, NME, MEN, and MNE). These six combinations of presentation order were examined with three different temporal intervals between face parts (0, 200, and 700 ms). If the order of part presentation is an important factor in temporal integration, we would expect to find it influencing mainly the 0p and 200p conditions. Specifically, the inversion effect in these two conditions should be higher in the ENM order than in the other five order options. In the 700p condition, this increased inversion effect in the ENM condition should be less emphasized due to the lack of temporal integration. However, if temporal integration can occur in spite of the presentation order of the parts, the inversion effect in the ENM condition should not differ from that in the other conditions. As in the previous experiment, the inversion effect should be more pronounced in 0p and 200p conditions than in the 700p condition across all the possible presentation orders.

Method

Participants. Sixty undergraduate students at the University of Toronto participated in the experiment for a course credit. All participants had normal or corrected-to-normal vision. None of them had participated in the previous experiments.

Materials. The same 180 famous faces of Experiment 2 were used in this experiment.

Procedure. Twenty participants were allocated for each of the three presentation-type conditions that were examined in the experiment (0p, 200p, and 700p). In contrast to Experiment 2, in which order of presentation was fixed (eyes part first on top, followed by the nose part in the middle, and finally the mouth part at the bottom), all the six possible combinations were investigated in Experiment 3. The face parts were always presented in their original position, but their order of presentation differed. Ninety faces were presented upright, and 90 faces were inverted. Within each orientation condition, 30 faces were presented in the original ENM order, and 12 faces were presented in each of the remaining five conditions. The ENM condition consisted of more trials than the other five conditions because we wanted to increase the probability of observing differences between conditions (if they indeed exist). If the order of presentation does influence temporal integration, then the inversion effect in the ENM condition should differ from that in the other five conditions, which could be collapsed together. The upright and inverted faces, as well as the different order conditions, were presented randomly during the experiment. The interval between the disappearance of the last part of the face and appearance of the mask was fixed, as in Experiment 2B (200 ms). Following the first part of the experiment, all faces were presented again as wholes for unlimited time in an upright condition, and participants were asked to recognize the faces. Twelve practice trials were administered prior to the experiment, representing all of the experimental conditions.

Results

The accuracy performance of each participant was computed, as before, as the proportion of the sum of correct responses in the first part of the experiment relative to the number of correct responses in the second part of the experiment, in which whole faces were presented upright for unlimited duration. A preliminary three-way ANOVA, with presentation type (0p, 200p, 700p) as a between-subjects factor and orientation (upright, inverted) and order of presentation (ENM, EMN, NEM, NME, MEN, MNE) as within-subject factors, performed on participants' mean accuracy in the second part of the experiment did not reveal any significant interactions between these variables. The mean accuracy of participants as a function of condition is presented in Table 3.

Our main aim in conducting Experiment 3 was to investigate whether temporal integration can be influenced by the order in which parts are presented. Specifically, will temporal interaction be affected when face parts are not presented in a serial consecutive order? The findings clearly show that temporal integration was successful even when the parts were not presented serially (see Figure 7). In each of the presentation type conditions, the inversion effect was comparable across the different presentation orders. This observation was corroborated in a three-way ANOVA, with presentation type (0p, 200p, 700p) as a between-subjects factor and orientation (upright, inverted) and order of presentation (ENM, EMN, NEM, NME, MEN, MNE) as within-subject factors, performed on participants' mean accuracy. The main effect of orientation was significant, $F(1, 57) = 533.79$, $MSE = 0.054$, $p < .0001$, resulting from greater accuracy to upright (.52) than to inverted faces (.11). In addition, a main effect of presentation type was found, $F(2, 57) = 4.80$, $MSE = 0.16$, $p < .01$. Finally, the Presentation Type × Orientation interaction was also significant, $F(2, 57) = 9.51$, $MSE = 0.054$, $p < .001$, indicating, as in the previous experiments, different inversion effects across the various interval conditions. Order of presentation was not significant either as a main effect, $F(2, 57) = 1.70$, $MSE = 0.017$, $p > .15$, or in interactions with other variables (all Fs < 1).

We conducted planned comparisons to examine the Presentation Type × Orientation interaction. As expected, comparable inversion effects were found for the 0p and 200p conditions ($F < 1$). The inversion effects in these two conditions were significantly greater than that found for the 700p condition, $F(1, 57) = 15.64$, $MSE = 0.054$, $p < .0005$, and $F(1, 57) = 12.87$, $MSE = 0.054$, $p < .001$, for the 0p and 200p conditions, respectively.

As in the previous experiment, we conducted additional post hoc analyses separately for upright and inverted faces to rule out a floor-effect interpretation of the reduction in the inversion effect. Again, no differences were found for inverted faces between the different interval conditions. However, for upright faces, the performance in the long-interval condition was significantly lower than that in the 0p and 200p conditions.

Discussion

The results of Experiment 3 provide an important replication of Experiment 2 by demonstrating again that the time course of temporal integration is approximately 450 ms, with comparable inversion effects appearing for temporal intervals less than 450 ms. More important, however, the present results show that within each specific condition of temporal interval between parts, similar inversion effects were obtained, regardless of the order in which face parts were presented. Thus, despite the fact that grouping by proximity of parts was not possible, a unified representation of a face was created.

The indifference of facial encoding processes to the presentation order of face parts, expressed by the comparable inversion effects, may indicate that configural processes are initiated only when all face parts are presented and maintained temporarily in a short-term visual buffer. Only when a whole image is stored in the visual buffer can advanced encoding processes, specific to faces, begin. According to this notion, configural processing is not part of temporal integration processes per se. Rather, temporal integration processes are a prerequisite for configural processing.

Table 3
Proportions of Correct Responses (With Standard Deviations in Parentheses) for Upright and Inverted Faces as a Function of Presentation Type and Order of Presentation in Experiment 3

	Order of presentation						
Presentation type	ENM	EMN	NEM	NME	MEN	MNE	Average
0p							
Upright	.63 (.21)	.57 (.23)	.61 (.25)	.54 (.22)	.58 (.19)	.62 (.25)	.59 (.16)
Inverted	.12 (.08)	.10 (.15)	.13 (.16)	.13 (.12)	.16 (.15)	.15 (.13)	.13 (.07)
Inversion effect	.50 (.20)	.47 (.25)	.49 (.24)	.41 (.21)	.43 (.17)	.47 (.27)	.46 (.15)
200p							
Upright	.53 (.13)	.51 (.21)	.61 (.16)	.56 (.21)	.49 (.23)	.55 (.22)	.55 (.13)
Inverted	.10 (.09)	.10 (.13)	.13 (.17)	.11 (.17)	.09 (.11)	.09 (.13)	.10 (.10)
Inversion effect	.44 (.11)	.41 (.20)	.48 (.20)	.45 (.17)	.40 (.17)	.46 (.23)	.44 (.09)
700p							
Upright	.43 (.16)	.37 (.24)	.47 (.25)	.44 (.17)	.42 (.23)	.44 (.24)	.40 (.19)
Inverted	.12 (.10)	.13 (.14)	.10 (.09)	.09 (.10)	.09 (.12)	.12 (.12)	.11 (.07)
Inversion effect	.31 (.13)	.24 (.24)	.38 (.25)	.35 (.17)	.33 (.25)	.31 (.22)	.29 (.14)

Note. The intervals between presentations of upright or inverted face parts were either 0 ms (0p), 200 ms (200p), or 700 ms (700p). E = eyes; N = nose; M = mouth.

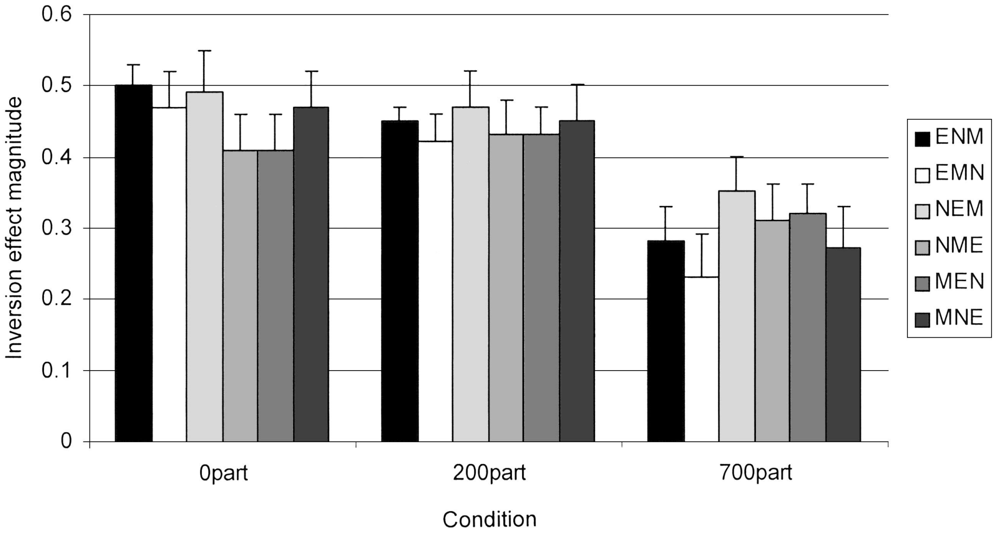

Figure 7. Magnitude of the inversion effects in Experiment 3 as a function of presentation type and order of presentation. Error bars represent standard errors. E = eyes; N = nose; M = mouth.

Alternatively, configural processes may be a crucial component of temporal integration processes while still being impervious to the order of part presentation. According to this view, facial configural processing involves the computation of spatial relations (or distances) between two (or more) internal features (e.g., Cooper & Wojan, 2000). Thus, the spatial relations between any two presented features can be encoded regardless of their proximity. This claim is especially relevant in the present study, in which face parts, although presented randomly, appeared at locations corresponding to their original positions. For example, the mouth part appeared at the lower third of the display even if it was presented first. Thus, participants were able to compute with great precision the distance between it and other features of the face. Our experiments indicate that a prerequisite for such configural processing is that all the elements co-occur within a limited period corresponding to the decay time of the visual buffer. Whether this conclusion applies only to the processing of faces or also extends to nonface objects has yet to be determined.

A related issue, which was not explored in the present study, is whether the location of the parts is an important factor in the integration process. In all of the experiments reported here, the face parts were presented in their correct position. Because of masking problems, the parts could not have been presented sequentially in the same location, but it is interesting to consider whether temporal integration could have been successful if the parts had been presented in a misaligned or scrambled format. Gauthier et al. (2006) reported configural effects at retrieval even when face parts were not presented at the canonical location for such parts. Thus, it may be possible to find temporal integration even when face parts are offset spatially and even when the top part appears at the bottom, and vice versa. If this is the case, it would argue for a VSTM locus for temporal integration, because evidence has shown that information in VSTM has abstracted some invariant features of the stimulus related to its identity and, thus, is not sensitive to spatial shift.

Misaligned presentation of the facial parts (as in Gauthier et al., 2006) may be used as an additional tool to pinpoint the locus of temporal integration: Because VSTM is not sensitive to spatial shift (Irwin, 1991), minimal effects should be seen when the to-be-integrated stimuli are offset. In contrast, representations in iconic memory are maintained in precise retinotopic coordinates, and a spatial offset, such as misalignment, would reduce performance. This procedure would possibly allow one to determine, in a similar fashion to Experiment 3, whether the integration in the intermediate-interval conditions (e.g., 200p) stems from integration in iconic memory or in VSTM.[4] However, it should be noted, that such an arrangement would require both temporal integration and spatial realignment. Although such a reconfiguration seems possible, it is also a time-consuming task that may not be achieved within the temporal boundaries delineated in our study.

Another related issue is the role of perceptual organization processes, such as grouping by proximity, in temporal integration. The fact that face identification was not influenced by part proximity, as evidenced by the fact that no difference was observed in the inversion effect between the ENM condition and the other conditions, may suggest that as with configural processing, grouping by proximity cannot occur early in perception or before a complete image is registered. This notion corresponds to traditional approaches to visual processing that posit that the perception and identification of global shapes are performed at a later stage in inferotemporal cortex regions, whereas in early visual areas only analysis of local features occurs (Felleman & Van Essen, 1991; Livingstone & Hubel, 1988). If this assumption is correct, the results of the current experiment point to iconic memory as the stage in which temporal integration takes place. Other findings also indicate that representations stored in iconic memory are not influenced by figural complexity (e.g., Irwin, 1991). Recent studies, however, suggest that grouping can occur early in visual processing and that brain activity related to grouping is observed in striatal and prestriatal regions (e.g., Altmann, Bülthoff, & Kourtzi, 2003; Han, Song, Ding, Yund, & Woods, 2001). The latter studies,

---

[4] We thank Andrew Hollingworth for this suggestion.

however, used simple geometric patterns or shapes, not complex ones such as faces. An important question for future research is whether the type of grouping that occurs among face parts has a similar locus in the perceptual stream and in the brain.

General Discussion

Temporal integration, in which time-segregated stimuli are combined to form a unified representation, has been demonstrated in the domains of object and scene perception. The findings of the present experiments establish that temporal integration can also occur in face perception. In the present study, famous and nonfamous faces were divided into three parts and presented sequentially, with a variable interval between each part. The influence of the temporal interval between the parts on the magnitude of the inversion effect was explored. Decrements in performance for upside-down faces are traditionally interpreted as resulting from disruption of holistic or configural processing and reliance on part-based, object-like processes. Thus, the presence or absence of an inversion effect is a sensitive marker, indicating whether a configural representation of the face was created through integration. We found that inversion effects were obtained for both famous and nonfamous faces even when face parts were integrated over a time span of 450 ms. Moreover, these inversion effects were similar to those found when a face was presented as a whole and not partitioned. In contrast, inversion effects were eliminated or significantly reduced when the delay between parts was 700 ms. A similar pattern was found with misaligned, unfamiliar faces, although it was less pronounced. Finally, temporal integration was not influenced by the order in which face parts were presented. Comparable inversion effects were obtained across the different orders of face part presentations within each presentation-type condition (i.e., the differing temporal intervals between the parts). Temporal integration was achieved even when nonadjacent parts were presented sequentially and when the mouth or the nose parts were presented first.

The theoretical framework we suggest to account for these findings is that the face parts are integrated in a time-limited visual buffer. Within this buffer, face parts are temporarily stored and combined with incoming additional face parts. The emergence of a face inversion effect indicates that the integration of the upright face parts is performed in a manner that allows the computation of the overall configuration of the original face; the relational information between the features is an essential component of the integration process. Because the inversion effect is considered a hallmark of configural processing, the emergence of an inversion effect in the temporal integration task in the present study strongly suggests that the facial features were not encoded in an isolated manner in the upright condition. Computation of the relational information failed, however, when face parts were presented upside-down, leading to impoverished identification.

The present findings strongly suggest that configural processing of faces is not dependent on simultaneous presentation of facial features. Rather, faces can still be perceived (configurally) if the different components of a face are separated by short temporal intervals. These findings support views claiming that face identification is dependent on the relational information between features that is computed during face processing (e.g., Diamond & Carey, 1986; Searcy & Bartlett, 1996).

The current findings could also be accommodated with accounts that regard face processing as involving template encoding, with little or no part differentiation (Farah, 2004; Farah et al., 1995, 1998; Tanaka & Farah, 2003), though some modifications to the original versions would be required. According to these accounts, although the component parts of the face are separable in principle, the perception and representation of the face is unparsed. One prediction, explicitly made by proponents of this approach, is that no holistic processing will occur when a face is *not* presented as a whole unit. As mentioned earlier, this hypothesis was investigated by Farah et al. (1995), who presented faces in a part-wise manner during a learning session, with each feature appearing in isolation or in a normal whole format. The existence of an inversion effect later in recognition in the whole condition, but not in the separated condition, was interpreted as evidence that a holistic representation cannot be formed when the parts are decomposed. Although we did not test for holistic processing directly, our findings of equivalent inversion effects for faces presented as wholes or in a piecemeal fashion suggest that holistic processing can occur even when the face parts are presented and perceived separately. Indeed, the temporal window during which temporal integration and, consequently, holistic processing can take place is narrow. This might explain Farah et al.'s (1995) failure to observe inversion effects in their study in which the presentation duration of each part was much longer. Yet, even within the limited temporal boundaries in our study, the face parts were explicitly encoded as decomposable elements, at least in the intermediate-interval conditions, but nevertheless were integrated in a holistic fashion.

Proponents of the template hypothesis could justly argue that holistic, template-like representations may be created *after* temporal integration processes are complete. The inversion effect may be attributed to that later stage. This proposal could easily account for the present results, but only if the original assumption—that during the perception of a gestalt, such as a face, "the whole stimulus takes precedence over the sum of its parts" (Tanaka & Farah, 2003, p. 53)—is relaxed. If the proponents of the template hypothesis would concede that holistic processing can arise even when separated facial parts are consciously perceived, they could easily interpret the current findings according to that hypothesis.

The locus of temporal integration processes in face perception has yet to be determined. As mentioned in the introduction, the integration of visual information across time intervals can take place in different memory systems (iconic memory, VSTM, or VLTM) and is heavily reliant on the nature of the representations that are formed in each of these memory systems. In the present context, because of the relatively short temporal intervals in which integration was observed, iconic memory and VSTM are the two most likely candidates. Iconic memory, although initially regarded as a unitary phenomenon, is actually fractionated into several subcomponents (Coltheart, 1980; Di Lollo & Dixon, 1988; Irwin & Yeomans, 1986). *Visible persistence* is the phenomenological perception of a visual trace that remains after stimulus offset. This trace is detailed and decays approximately 80–100 ms after the onset of the stimulus. *Informational persistence* is a second, more durable mechanism, which is time-locked to stimulus offset and decays within 150–300 ms. Although informational persistence contains elaborate form and spatial information, it is not visible directly and is not accompanied by the same perceptual experience that characterizes visible persistence. In contrast to iconic memory, VSTM is able to maintain abstracted information across several seconds (Phillips, 1974), but it has a limited capacity of 3–4 objects (Luck & Vogel, 1997), and its spatial information is less accurate. Findings have shown that visible persistence and VSTM could support integration, but informational persistence could not (e.g., Di Lollo, 1980; Irwin, 1993, 1996).

Because the temporal window of facial-features integration overlaps, to a large degree, the boundaries of iconic memory, we are inclined to conjecture that the processes involved in creating a unified face more likely occurred within iconic memory than within VSTM. If so, why was temporal integration successful even at the 200p condition, in which approximately 450 ms separated the first and third facial parts? Although temporal integration observed at the shorter intervals could be attributed to visible persistence, the integration observed in the 200p condition is seemingly within the time frame of informational persistence, which supposedly does not support integration.

One simple explanation for this divergence may be related to the different paradigms used in the present study as compared with previous studies. Whereas in studies in which no integration was found, the stimuli were spatially overlapping, in the present paradigm they were not. This raises the possibility that integration could be established during informational persistence as well, but only when the stimuli do not overlap and no backward masking mechanism is functioning.[5]

Additionally, facial integration could be based on informational persistence because of the unique nature of faces, for which human observers' lifelong expertise has allowed fine-grained and highly sophisticated mechanisms to develop. Thus, the computation of spatial information when faces are presented may be more proficient and accurate than that when other types of stimuli are presented in tasks that require fine visual alignment (Di Lollo & Dixon, 1988). As a result, the dissipation of the memory trace may be decelerated when faces are concerned, or, alternatively, significant information could be extracted despite the rapid decay.

Although this conjecture is admittedly speculative, some support for it can be found in a recent study by Hollingworth et al. (2005). They used the empty-cell localization task, in which two arrays with dot patterns are superimposed, with varying intervals between the arrays, and the participant is asked to report the square that was not filled with a dot. Varying the complexity of the first array, Hollingworth et al. found less accuracy for the complex array than for the simple array at an interarray interval of 100 ms. Although their main interest and predictions were focused on longer intervals, this result supports our speculation that even at the level of informational persistence, complexity may play an important role. As a result, the expertise of the perceiver with the visual stimuli will facilitate temporal integration.

The diminution, or even complete elimination, of the inversion effect in the long-interval condition negates, in our view, the possibility that facial integration could occur in VSTM. The inability to integrate face parts in VSTM could be a result either of capacity limitations, which limit its span to 4–5 items (Cowan, 2001; Sperling, 1960), or inability of VSTM to maintain accurate spatial information (Hollingworth et al., 2005; Irwin, 1991; Phillips, 1974). Recent results, however, reported by Brockmole, Wang, and Irwin (2002; Brockmole & Wang, 2003), have questioned the traditional view of VSTM as a limited-capacity system by demonstrating an average memory span of approximately 10 items, which exceeds by more than twofold the accepted view of VSTM capacity. Brockmole and colleagues suggested that integration in VSTM is between image and percept, where the first stimulus is transformed into an image and maintained in a visual buffer until it is combined with the second stimulus (Kosslyn, 1994). This pattern of integration occurs only when the interval between the two stimuli is long enough to allow the formation of an internal representation of the first stimulus. This finding was replicated recently by Hollingworth et al. (2005). We did not observe such a pattern with faces in that no integration occurred at the longest interval.

The reason for these discrepancies may stem both from the nature of the representations created in VSTM and the characteristics of the tasks and stimuli used in the different studies. The memory traces maintained by VSTM are abstract, postcategorical representations that code the visual, but not the semantic, features of the stimulus (Henderson & Hollingworth, 2003; Hollingworth, 2004). As such, they are influenced by configural grouping of single elements into more high-order configurations, and chunking could maximize the capacity of VSTM, as demonstrated with dot patterns (Hollingworth et al., 2005). Faces, however, differ from dot patterns, and the issue of how faces are represented in VSTM, and whether representations of faces in VSTM differ from long-term memory representations, has yet to be addressed (Cooper & Wojan, 2000; Diamond & Carey, 1986; Rhodes, 1988). In addition, it is important to note a crucial difference between the empty-cell localization task and the task used in our study. In the latter task, construction of a structural and fixed abstract description of a face was not possible until the entire face was presented. For example, the exact position of the eyes in the face was determined only when all of the parts were presented, because identification is based on the configural relations between the features. In contrast, in tasks in which dot-filled matrices are presented, a reference frame can be established for each array separately, because the location of the dots can be drawn in relation to the matrix, which is constant across arrays. This greatly facilitates the ability to represent abstractly the arrays in VSTM, but it makes it difficult to apply such long-lived, part-based representation to faces for the purpose of integrating the parts into a whole.

In short, because faces differ on many dimensions from the other stimuli that have been used to study VSTM—as do the procedures in our study from those used to study VSTM for dots, objects, and scenes—it is difficult to extrapolate easily from the conclusions of other studies to our own. Nonetheless, insofar as it is possible to make such comparisons, our results suggest that the locus of temporal integration for faces is not in VSTM.

The findings of the present study emphasize the importance of a temporally integrated representation for configural processing of faces. Integrated faces will not necessarily be perceived in a configural manner, but, nevertheless, configural perception of faces necessitates visual integration. Our data show that this integration is still apparent at approximately 400 ms, which favors iconic memory as the platform supporting temporal integration but does not completely rule out the possibility that the locus is in VSTM.

---

[5] We thank Andrew Hollingworth for this suggestion.

This theoretical account may have some interesting (although speculative) implications for the neuropsychological literature in that it suggests that some forms of prosopagnosia may arise either from the disruption of the integrative processes or from damage to the visual buffer itself. The possibility of a faulty integrative mechanism underlying prosopagnosia seems to contradict the double dissociation observed between deficits in face and object perception, both of which, allegedly, depend on temporal integration. However, as already claimed in the literature, these integrative processes may not be similar across domains, and they may differ qualitatively from one another due to factors such as expertise and complexity (e.g., Gauthier & Nelson, 2001; Kanwisher, 2000; Maurer et al., 2002; Moscovitch et al., 1997). The possibility that deficits in face perception (and object perception as well) may stem from damage to the visual buffer or its functioning has not been proposed, to our knowledge, but may prove a promising venue of future investigation.

References

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology, 13,* 342–349.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281–316.

Barton, J. J. S., Deepak, S., & Malik, N. (2003). Attending to faces: Change detection, familiarization, and inversion effects. *Perception, 32,* 15–28.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In W. Waggenspack (Chair), *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 187–194). New York: ACM Press/Addison-Wesley.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Bridgeman, B., van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences, 17,* 247–292.

Brockmole, J. R., & Wang, R. F. (2003). Integrating visual images and visual percepts across time and space. *Visual Cognition, 10,* 853–873.

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 315–334.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & Psychophysics, 65,* 1296–1306.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Cooper, E. E., & Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 470–488.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114.

Davies, T. N., & Hoffman, D. D. (2002). Attention to faces: A change-blindness study. *Perception, 31,* 1123–1146.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75–97.

Di Lollo, V., & Dixon, P. (1988). Two forms of persistence in visual information processing. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 671–681.

Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology, 26,* 33–63.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4,* 345–352.

Farah, M. J. (2004). *Visual agnosia* (2nd ed.). Cambridge, MA: MIT Press.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628–634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498.

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences, 8,* 265–289.

Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception, 19,* 661–673.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11,* 349–370.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology, 11,* 219–224.

Gauthier, I., Tanaka, J. W., & Brown, D. D. (2006). Are configural effects in face processing due to holistic encoding? *Manuscript submitted for publication.*

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446.

Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception, 15,* 235–247.

Han, S., Song, Y., Ding, Y., Yund, E. W., & Woods, D. L. (2001). Neural substrates for visual perceptual grouping in humans. *Psychophysiology, 38,* 926–935.

Heider, B., & Groner, R. (1997). Backward masking of words and faces: Evidence for different processing speeds in the hemispheres? *Neuropsychologia, 35,* 1113–1120.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 356–383). New York: Oxford University Press.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113–136.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 519–537.

Hollingworth, A., Hyun, J. S., & Zhang, W. (2005). The role of visual short-term memory in empty cell localization. *Perception & Psychophysics, 67,* 1332–1343.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761–768.

Humphreys, G. W., Hodsoll, J., & Campbell, C. (2005). Attending but not seeing: The "other race" effect in face and person perception studied through change blindness. *Visual Cognition, 12,* 249–262.

Ikeda, M., & Uchikawa, K. (1978). Integrating time for visual pattern perception and a comparison with the tactile mode. *Vision Research, 18,* 1565–1571.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23,* 420–456.

Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in* *experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 121–142). Cambridge, MA: MIT Press.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5,* 94–100.

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 343–360.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64,* 882–895.

Jonides, J., Irwin, D. E., & Yantis, S. (1982, January 8). Integrating visual information from successive fixations. *Science, 215,* 192–194.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3,* 759–763.

Kanwisher, N., & Moscovitch, M. (2000). The cognitive neuroscience of face processing: An introduction. *Cognitive Neuropsychology, 17,* 1–13.

Keppel, G. (1982). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112,* 24–38.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge MA: MIT Press.

Kubovy, M., Holocombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35,* 71–98.

Kubovy, M., & Pomerantz, J. R. (Eds.). (1981). *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*(A), 449–473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*(A), 513–536.

Leder, H., Candrian, G., Hubber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30,* 63–73.

Livingstone, M., & Hubel, D. (1988, May 6). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740–749.

Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research, 45,* 2287–2297.

Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology, 35,* 135–199.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260.

McConkie, G., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 563–581.

McConkie, G., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (2nd ed., pp. 137–162). Newark, DE: International Reading Association.

McKone, E., Martini, P., & Nakayama, K. (2003). Isolating holistic processing in faces (and perhaps objects). In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 92–119). New York: Oxford University Press.

Moscovitch, M., & Moscovitch, D. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology, 17,* 201–219.

Moscovitch, M., & Radzins, M. (1987). Backward masking of lateralized faces by noise, pattern, and spatial frequency. *Brain and Cognition, 6,* 72–90.

Moscovitch, M., Winocour, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9,* 555–604.

Murray, J. E. (2004). The ups and downs of face perception: Evidence for holistic encoding of upright and inverted faces. *Perception, 33,* 387–398.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461–488.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24,* 939–1011.

Palermo, R., & Rhodes, G. (2003). Change detection in the flicker paradigm: Do faces have an advantage? *Visual Cognition, 10,* 683–713.

Peterson, M. A., & Rhodes, G. (2003). Introduction: Analytic and holistic processing—The view through different lenses. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 3–19). New York: Oxford University Press.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290.

Psychological Software Tools. (2000). E-prime (Version 1.0b) [Computer software]. Pittsburgh, PA: Author.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245–277.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17,* 43–63.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47,* 25–57.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12,* 94–99.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 257,* 9–15.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904–915.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Simons, D. J., Mitroff, S. R., & Franconeri, S. L. (2003). Scene perception: What we can learn from visual integration and change detection. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 335–355). New York: Oxford University Press.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9,* 16–20.

Singer, J. M., & Sheinberg, D. L. (2006). Holistic processing unites face parts across time. *Vision Research, 46,* 1838–1847.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498).

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 53–74). New York: Oxford University Press.

Tanaka, J. W., & Sengco, J. A. (1997). Faces and their configuration in face recognition. *Memory & Cognition, 25,* 583–592.

Trehub, A. (1977). Neuronal models for cognitive processes: Networks for learning, perception, and imagination. *Journal of Theoretical Biology, 65,* 141–169.

Troje, N., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36,* 1761–1771.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471–491.

Veuilleumier, P., Mohr, C., Valenza, N., Wetzel, C., & Landis, T. (2003). Hyperfamiliarity for unknown faces after left lateral temporo-occipital venous infarction: A double dissociation with prosopagnosia. *Brain, 126,* 889–907.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6,* 313–326.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98,* 4800–4804.

Wertheimer, M. (1925). Concerning Gestalt theory. *Symposion, 1,* 39–60.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Young, A. W., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception, 10,* 747–759.

Yovel, G., Paller, K. A., Levy, J. (2005). A whole face is more than the sum of its halves: Interactive processing in face perception. *Visual Cognition, 12,* 337–352.

Received July 2, 2005
Revision received April 9, 2006
Accepted April 19, 2006 ■

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.

View publication stats