Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 46 (2006) 1838–1847
Vision Research
www.elsevier.com/locate/visres

Holistic processing unites face parts across time

J.M. Singer [a], D.L. Sheinberg [b,*]

[a] Brain Sciences Program, Brown University, Box 1953, Providence, RI 02912, USA
[b] Department of Neuroscience, Brown University, Box 1953, Providence, RI 02912, USA

Received 18 July 2005; received in revised form 1 November 2005

Abstract

When complementary halves of different familiar faces are combined into a new face, there is interference in the identification of either half. This "composite face effect" has been taken as strong evidence that faces are processed holistically. Here, we demonstrate that this effect can persist when the two parts of a face are separated by up to 80 ms of visual noise, showing that the parts of a face interact not only spatially but also temporally. We suggest that the processing underlying robust identification accepts an accumulation of evidence over time.

Keywords: Face perception; Configural processing; Holistic processing; Object recognition; Temporal integration

1. Introduction

The ability to rapidly and accurately identify faces is critical to normal social function. Yet we know remarkably little about how this important and everyday process works. Even the most advanced techniques for automated computer recognition of faces are successful only under highly constrained circumstances (Kong, Heo, Abidi, Paik, & Abidi, 2005), and there is little consensus about the methods by which our visual system solves this problem.

Galton (1879) pointed out that faces are generally very similar to one another; the corollary is that if we are to identify them by other than a minute and painstaking inspection we must somehow rapidly extract information from the whole face. He made this explicit soon after (Galton, 1883): "… a face is the sum of a multitude of small details, which are viewed in such rapid succession that we seem to perceive them all at a single glance." Since then, there have been many models of face perception incorporating some sort of "holistic" processing that allows for our remarkable abilities. In their review, Farah, Wilson, Drain, and Tanaka (1998) argue for a particular sort of holistic processing in which faces are broken down into component parts to a much lesser degree than other objects. We might recognize our car in a parking lot by searching for the right color, the right hubcaps, the right license plate, and the right radiator grille, but we recognize Bob's face in a crowd much more atomically, simply as Bob's face.

This is by no means the only denotation holistic has had. Gauthier and Tarr (2002) describe three sorts of holistic face processing. A part of a face is better recognized when all the parts are in their proper locations relative to each other; this effect they call *holistic-configural*. They describe *holistic-inclusive* as the influence on a part's identification played by the identities associated with all the nearby parts. Finally, *holistic-contextual* effects enable a part to be recognized better in context than in isolation. "Configural" is another term that has frequently been used to describe the processing of faces as more than just collections of independent parts. Its usage has been similar to, and in many cases synonymous with, holistic. Maurer, Le Grand, and Mondloch (2002) divide configural processing into three types; first-order (the nose is above the mouth) and second-order (the eyes are separated by 5.5 cm) spatial relations, and holistic processing as described by Farah et al. (1998).

* Corresponding author. Tel.: +1 401 863 9575.
E-mail addresses: Jedediah_Singer@brown.edu (J.M. Singer), David_Sheinberg@brown.edu (D.L. Sheinberg).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.11.005

While the details and nomenclature are still under discussion, there is consensus that faces are processed as more than a linear function of their component parts. When this processing is disrupted, robust identification suffers.

One of the strongest pieces of evidence for holistic processing is the "composite face effect." Young, Hellawell, and Hay (1987) found that it is inordinately difficult to identify a specified part of a composite face, created by joining half of one face to the complementary half of another face. The difficulty in recognizing these incongruent chimeras relative to normal (congruent) faces is ameliorated if the composite is turned upside down or the two halves are misaligned. This implies that all the parts of a properly configured face inevitably interact; when those parts indicate different identities, identification is necessarily impeded. When the configuration of the face is disrupted by inversion or misalignment, however, these interactions are in large part abolished.

To date, there has been very little investigation into the dynamics of face recognition itself—i.e., how the determination of a face's identity develops across time—and none at all about the time course of holistic identity processing. Yet face identification does *not* occur "in a flash"; there is a delay, often imperceptible but sometimes quite considerable, between the time that a face is detected and the time that it is identified. Work with moving faces (Lander & Bruce, 2000) suggests that this delay is not due to ongoing processing of a static snapshot, but rather to an accumulation of evidence. As described above, even Galton referred to the "rapid succession" of details that constitute what appears to be a single glance.

Here, we report two experiments that examine the time course of holistic face processing. In particular, we ask if parts of a face interact with other parts that are present at different times. If so, how large a temporal gap can this influence span? One might expect that synergistic interactions such as those seen in the composite face effect require the parts to be present at precisely the same time. Alternatively, information from sequentially presented parts may accumulate and interact, giving rise to a composite face effect even when there is not a whole face present at once. We predict that the latter will hold true, as long as the first part presented is sufficient to trigger holistic processing. We used a modification of the procedure from Young et al. (1987), and examined the composite face effect when face halves were presented at varying times relative to one another.

2. Experiment 1

We tested for a composite face effect using briefly presented famous faces consisting of separable interior and exterior regions. The target part of the face (the interior) could appear at the same time as the irrelevant part (the exterior), or with a stimulus onset asynchrony (SOA) of up to 200 ms in either direction.

2.1. Method

2.1.1. Stimuli

Original faces were grayscale images of Bill Clinton and John Kerry, four of each, downloaded from publicly accessible web sites. Faces were cut out of their backgrounds above the neck and intensity histograms were then equalized using Adobe Photoshop CS (Adobe Systems Inc., San Jose, CA). The same alpha selection mask (up to translation) was used for all faces to separate them into inner and outer halves (centers and surrounds), with a several-pixel feathered region in-between that was identical across all faces.

Noise was generated by phase-scrambling the average spectrogram of the eight original faces. The resulting noise textures subtended approximately 6° of visual angle, and the embedded faces nearly that much top-to-bottom. Stimuli were presented on a CRT monitor with refresh rate of 100 Hz at a distance of approximately 100 cm.

Each trial began with the onset of a fixation point and a tone, followed by a blank screen, and then a sequence of noise frames that changed every 40 ms (25 Hz). One face half was superimposed on one randomly selected noise frame in the interval 120–400 ms, and the other face half was either in that same frame or in a frame 40–200 ms later (Fig. 1). SOA was defined as the time between the onset of the target interior and the onset of the irrelevant surround, and was considered negative if the interior followed the surround. Face halves were either both upright or both inverted, and could be from the same face (congruent condition) or from faces of different identity (incongruent condition).

2.1.2. Procedure

Subjects were familiarized with the stimuli in a practice block composed of the original four images shown both upright and inverted eight times each, in random order. There were 1600 test trials, randomly shuffled into blocks of 100. Congruent faces were repeated four times to match the number of incongruent trials. This accounts for 1440 trials (11 SOAs × 2 conguencies × 2 orientations × 4 repetitions × 8 face images); the other 160 trials were either isolated centers or hybrids in which the orientation of the center differed from the orientation of the surround. These 160 trials were not analyzed for the present experiment.

Subjects indicated the identity of the face center by pressing one of two buttons. Accuracy was repeatedly emphasized; subjects were instructed to answer as quickly as they could, but only once they were sure they knew the proper identity. They were told to guess if a few seconds had passed and they were still unsure. Feedback was provided by an auditory cue in both correct and incorrect trials, and trials were terminated and counted as incorrect if no response was given within 5 s.

2.1.3. Participants

Sixteen naïve subjects (7 male) aged 18–30 years participated after signing a consent form. Data from one subject

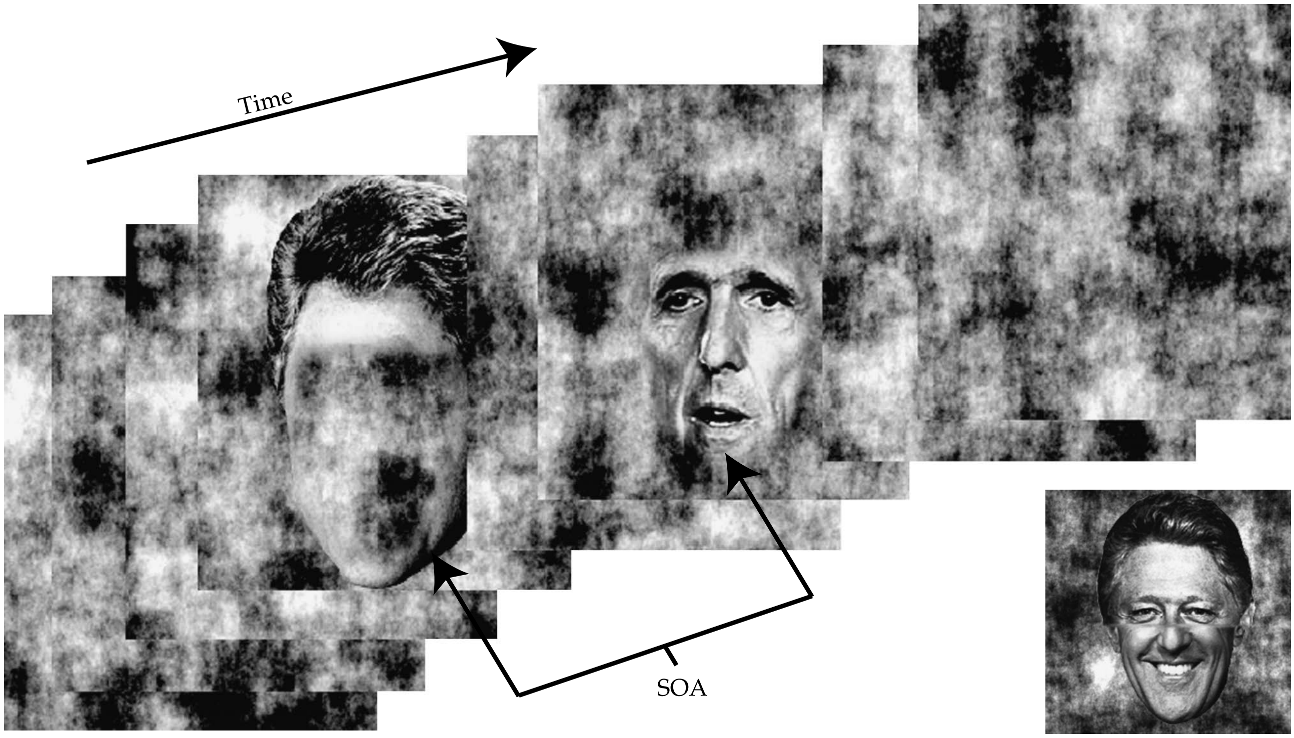

Fig. 1. Inner and outer face parts were embedded in two 40 ms frames of a dynamic stream of noise; subjects were required to indicate the identity of the inner part. SOA ranged in one-frame (40 ms) steps from −200 to 200 ms. In this example, SOA was −80 ms. The inset shows an example of a top/bottom incongruent stimulus from Experiment 2.

were discarded due to below chance performance on 7 of 44 conditions (SOA × orientation × congruence). Subjects were undergraduates, graduate students, and professionals, and were compensated for their time. All recruitment and testing were in accordance with procedures approved by the Brown University IRB.

2.2. Results

We examined correct reaction times, calculated from the onset of the target part of the face, for each subject. We also discarded correct trials at both reaction time extremes, when it seemed likely that the responses corresponded to uninformed guessing or unintentional button presses (indicated by below-chance performance in a 200 ms bin). In total, these trimmed extreme trials amounted to 0.87% of correct trials. Each subject's mean response times were then computed for each of the 44 conditions, and the results were averaged across subjects.

In general, inversion and incongruence both slowed responses. Population reaction times were significantly faster in each of the four conditions (ANOVAs, all $p < 0.005$) when the irrelevant exterior preceded the interior target (negative SOAs), though the shapes of the curves differ. Congruent trial reaction times increased monotonically as the exterior appeared increasingly later than the interior, while the incongruent trial reaction times followed a more peaked trajectory. Fig. 2 displays these raw reaction-time results. Upwards-pointing triangles represent upright conditions, and black triangles represent incongruent conditions. Error bars indicate standard errors of the mean.

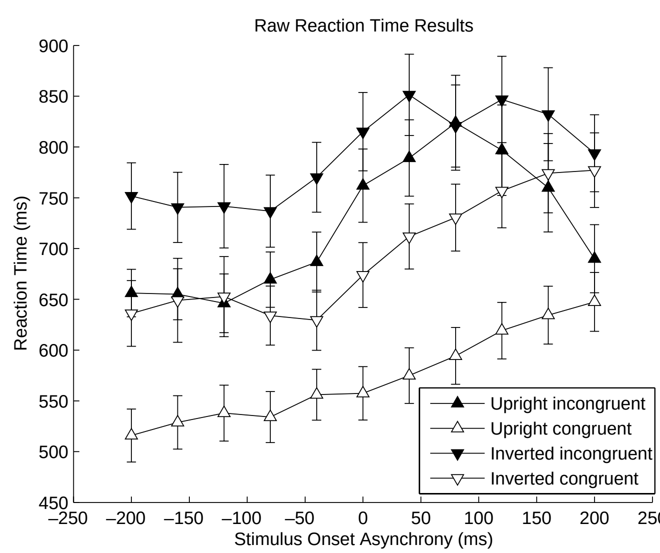

Fig. 2. Abscissa is SOA, ordinate is reaction time averaged across all subjects. Upwards-pointing triangles indicate upright trials, and unfilled and filled triangles represent congruent and incongruent trials, respectively. Error bars indicate standard errors of the mean. Reaction times are faster when the irrelevant exterior precedes the interior target; congruent trials slow monotonically with increasing SOA, while incongruent reaction times peak and then begin to improve at long SOAs.

We subtracted the mean and divided by the standard deviation of the reaction times for each subject, so that the data for each subject had a mean of zero and a variance of one. These normalized data were used in a three-way ANOVA (SOA × orientation × congruence). There were significant main effects of SOA ($F = 46.86$, $p < 0.0001$), orientation ($F = 441.46$, $p < 0.0001$), and congruence ($F = 835.97$, $p < 0.0001$); interactions were significant between orientation and congruence ($F = 28.12$, $p < 0.0001$) and between SOA and congruence ($F = 8.83$, $p < 0.0001$), but not between SOA and orientation ($F = 0.8$, $p > 0.6$) or between all three factors ($F = 1.65$, $p > 0.05$).

The magnitude of the composite face effect at a given SOA was calculated as a difference of differences. We subtracted the interference when inverted—the reaction time difference between congruent and incongruent inverted faces—from the interference when upright. This gives a positive value when the upright reaction times are more affected by incongruent surrounds than are the inverted reaction times, and thereby yields a numerical value for the composite face effect. Fig. 3 shows these results at all 11 SOAs. The classical composite face effect, greater interference from an irrelevant face-half when upright than when inverted, is replicated as a 63 ms reaction time difference at 0 ms SOA. More strikingly, the magnitude of the effect grows with increasing asynchrony (irrelevant surround following target) before dropping just below significance at an SOA of 160 ms. In other words, two parts of a face separated by 80 ms of uninformative noise (SOA of 120 ms) interact and give rise to a significant composite face effect.

Significance at each SOA was determined by a complete permutation test, which makes no assumptions about the distribution of the data. The hypothesis tested was zero composite face effect—in other words, the strength of the interference (congruent minus incongruent reaction times) was no different between upright and inverted conditions. If this were true, negating the values of the composite face effect magnitudes (i.e., swapping upright and inverted) for any subset of subjects would yield another data point equally likely as the one observed. With 15 subjects there are 2^{15} such data points, which give an empirical approximation of the hypothetical distribution about zero. When the actual data point is close to the edge of this distribution, it casts doubt on the null hypothesis. The data in Fig. 3 are plotted along with the 95th, 99th, and 99.9th percentiles of this distribution. A one-way ANOVA confirms that the composite face effect is significantly greater at positive SOAs than at negative ($F = 10.21$, $p < 0.005$).

The average discriminability across all subjects for each condition, measured by A' (Pollack & Norman, 1964; Macmillan & Creelman, 1996), is plotted in Fig. 4. Again, upwards-pointing triangles represent upright conditions, black triangles represent incongruent conditions, and error bars indicate standard errors. Discriminability was nearly perfect in upright congruent conditions, and generally above 0.95 when congruent but inverted. Upright incongruent trials as a group had an average A' of 0.92, but showed a marked dip in discriminability centered around 80 ms SOA. Performance on inverted incongruent trials varied widely between subjects and was on average much worse than the other three conditions.

To analyze the performance data, A' measures for each subject were normalized to zero mean and unit variance as for reaction time, and used in a three-way ANOVA (SOA × orientation × congruence). There were significant main effects of orientation ($F = 78.67$, $p < 0.0001$), congruence ($F = 267.64$, $p < 0.0001$), and SOA ($F = 2.37$, $p < 0.01$).

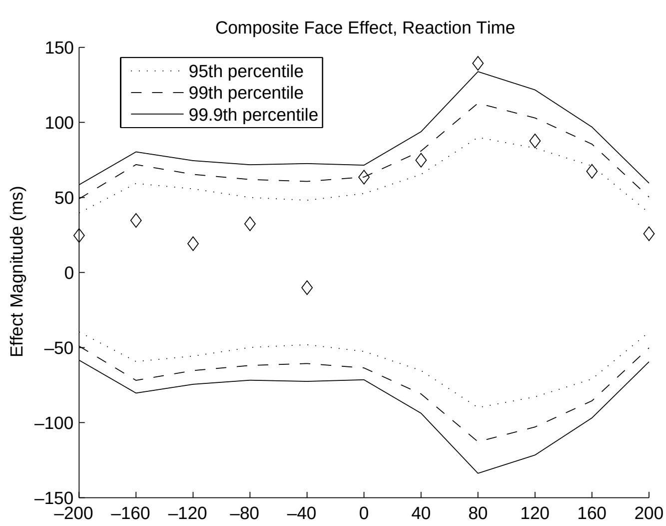

Fig. 3. Abscissa is SOA, ordinate is the magnitude of the reaction time composite face effect. Dotted, dashed, and solid lines indicate respectively the 95th, 99th, and 99.9th percentiles given by a complete permutation test. The composite face effect is significant from 0 to 120 ms SOA, and drops just below the 95th percentile at 160 ms SOA.

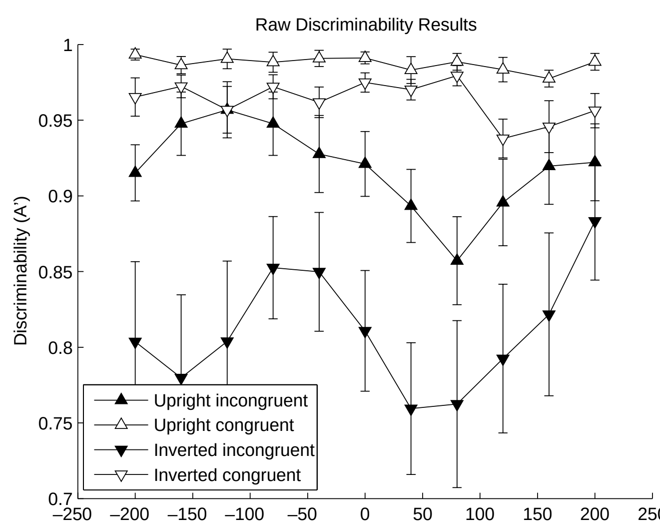

Fig. 4. Abscissa is SOA, ordinate is A' averaged across all subjects. Upwards-pointing triangles indicate upright trials, and black triangles represent incongruent trials. Error bars indicate standard errors. Performance suffers in incongruent trials, and dramatically when stimuli are also inverted. There is a dip in upright incongruent performance centered around 80 ms SOA, the asynchrony at which upright incongruent reaction times were slowest.

The only significant interactions were between orientation and congruence ($F = 23.34$, $p < 0.0001$) and between SOA and congruence ($F = 2.74$, $p < 0.005$). Casual inspection of these data in Fig. 4 suggests that there is no composite face effect with respect to discriminability. In fact, there is a negative effect at all SOAs—in other words, there is arithmetically *more* interference from an incongruent surround when the faces are inverted. Note though, that subjects performed near ceiling in the upright congruent case, making analysis of the difference measure used for reaction times unconvincing.

The dangers of examining a difference of differences without appropriate data were made explicit by Loftus, Oberg, and Dillon (2004). Ceiling and floor effects can make the magnitudes of the two differences meaningless; more generally, any nonlinearity in the function that maps the independent variables to the dependent variables can do the same. A particular change in one variable (e.g., inversion) might result in small behavioral changes when the task is otherwise easy (e.g., congruent surrounds) but large changes when it is difficult. This might possibly indicate interesting differences in the way faces are processed under different circumstances; it might simply be, however, that changes in task difficulty affect performance to a varying degree depending on whether the task is easy or hard.

To ensure that the reaction time results are not so trivially explained away, we follow the advice of Loftus et al. (2004) and turn to "dimensional theory." This posits that the observed dependent variables are constrained functions of some underlying psychological dimension or dimensions, which in turn are functions of the independent variables. A monotonic state-trace plot (a plot in which dependent variables are plotted against each other) indicates that the results can be explained by a single underlying psychological dimension, for example, "stimulus strength" or "task difficulty." The inverse is also true: if the state-trace plot is not monotonic, a one-dimensional model is insufficient to describe the data. We would thus expect to see a non-monotonic state-trace plot if an incongruent surround (holistically) affects upright faces more than inverted faces, rather than simply making the task more difficult across the board.

Fig. 5 shows the state-trace plot for reaction times. Circles indicate congruent trials, and crosses signify incongruent trials. For a given SOA and congruence, the average upright reaction time across all subjects is plotted against the average inverted reaction time. This yields 11 data points, corresponding to the eleven SOAs, for each congruence condition. Both congruent and incongruent conditions show approximately linear relationships (R^2 of 0.92 and 0.85, respectively) between inverted reaction times and upright reaction times, indicating that SOAs that were slower at one orientation tended to be slower at the other. An ANOCOVA reveals that the slope of the best-fit line through the congruent data points is significantly lower than for the incongruent data points ($F = 10.84$, $p < 0.005$). Moreover, the upright reaction

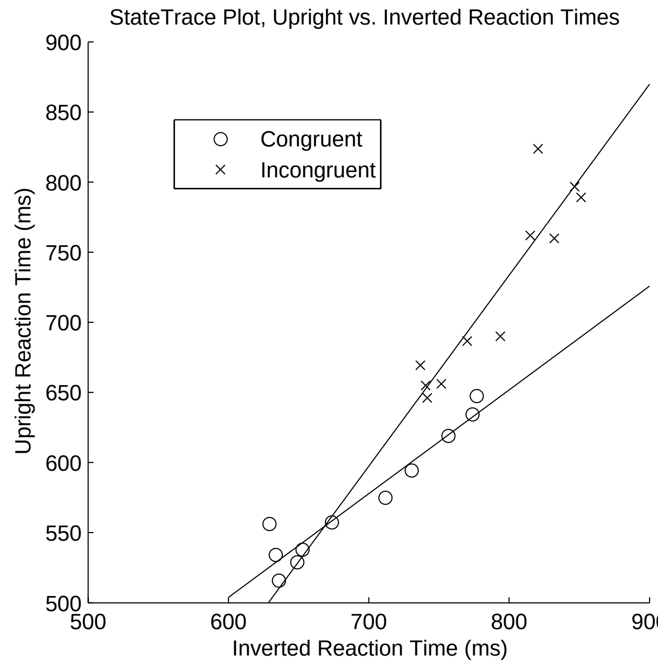

Fig. 5. State-trace plot of upright reaction times against inverted reaction times. Each data point plots the average reaction time for upright trials of a particular SOA and congruency versus the average reaction time for inverted trials of the same SOA and congruency. Congruent trials are indicated by circles, incongruent trials by crosses. For each congruency class, the best-fit lines through the data are plotted. The lack of monotonicity indicates that there must be more than one psychological dimension responsible for the reaction times, at least one of which is affected differently by upright faces than by inverted faces.

times for congruent trials are significantly lower than for incongruent trials, even when considering only those inverted reaction times (735–780 ms) in which the two congruency conditions overlap (ANOVA, $F = 19.78$, $p < 0.005$). This suggests that the two curves really are different, and therefore the results cannot be attributed solely to variation along a single psychological dimension (such as stimulus strength).

2.3. Discussion

Previous studies (Boutet, Gentes-Hawn, & Chaudhuri, 2002; Gauthier, Tanaka, & Brown, 2003; Hole, 1994; Young et al., 1987; Yovel, Paller, & Levy, 2005) have reported composite face effects under several sets of conditions. The present experiment extends these findings by showing that the composite face effect is found for briefly presented faces split inside/outside, with inversion used to disrupt holistic processing. More significantly, our results demonstrate the composite face effect when the irrelevant outside half of the face follows the target interior half by over 100 ms.

The composite face effect found in reaction times arises from slower responses in the upright incongruent cases at the affected SOAs. Faster reaction times in the upright congruent trials might have indicated some facilitation of those trials. If the effect were due to variation in the inverted trials, it would be more difficult to argue for the importance of interference in the upright incongruent cases. The effect does not appear to be due to a trivial speed/accuracy tradeoff, either, as upright incongruent performance of the population is actually worse at SOAs of 0–160 ms (ANOVA, $F = 9.04$, $p < 0.015$). The raw data suggest that the effect of holistic processing tapers off as SOA increases past 120 ms: reaction times in upright incongruent cases improve dramatically, and by 200 ms SOA there is no significant difference between congruent and incongruent faces. We will return to the difference between negative SOAs (distractor preceding target) and positive SOAs (distractor following target) in Section 4.

State-trace analysis reveals that no one psychological dimension can account for the pattern of reaction times. Considering inverted reaction times, results at the fastest incongruent SOAs were faster than the slowest congruent SOAs. When comparing upright reaction times at these same SOAs, however, the incongruent trials were slower (see Fig. 5). For this to happen, these two independent variables must influence reaction time through at least two underlying mechanisms. This demonstrates that the composite face effect measured was not due to a simple nonlinearity in, for example, the function mapping difficulty to reaction time. It appears to truly reflect two different modes of processing.

Subjects were near perfect accuracy on upright congruent faces, but many found the inverted incongruent trials extremely difficult. It is possible that a difference-of-differences composite face effect would manifest in the discriminability results if performance were not constrained at one or both ends of the scale. However, given the temporal limitations intrinsic to the experiment, the wide range in difficulty across conditions, and the need for trials to be matched with regards to non-experimental parameters (for example, contrast), appropriately adjusting difficulty was not feasible in this experiment. Past experiments have revealed composite face effects evident in reaction time (Hole, 1994; Young et al., 1987), in performance (Yovel et al., 2005), in both (Gauthier et al., 2003), and in neither (Hole, 1994), depending on the design of the experiment.

3. Experiment 2

Much of the previous work with composite faces has dealt with faces split between top and bottom rather than inside/outside. We therefore ran a second experiment, using horizontally split faces, with three questions in mind. First, would we see holistic effects across temporal gaps with these horizontally divided faces? Second, would inversion and misalignment disrupt the same processes, or would their effects be additive? Finally, why did we see such a large effect when the irrelevant half followed the target but not vice versa?

3.1. Method

Fifteen subjects were run (5 male, aged 18–43 years), five of whom had previously participated in Experiment 1. The same eight faces were split horizontally just below the eyes (see Fig. 1 inset). Subjects reported the identity of the top face half. Only SOAs of −80, 0, and 80 ms were used, to minimize the total number of trials. These three SOAs were chosen as representative of the target-first, target-second, and simultaneous conditions.

In addition to and independently from inversion, the halves could appear either properly aligned or horizontally offset by 1.5°. This yielded 764 trials (3 SOAs × 2 congruencies × 2 orientations × 2 alignments × 4 repetitions × 8 face images) which were randomly shuffled into eight blocks of 96. A preliminary training block of 96 trials, randomly selected from the 764, was presented prior to the experiment. There was no tone at the start of a trial, and auditory feedback was given only in the event of a mistake (this change was made at the suggestion of several subjects in the first experiment, who felt that the repetitive tones lulled them into a rhythm that was harmful to accuracy on difficult trials). In all other ways this experiment was identical to Experiment 1.

3.2. Results

Only reaction time data will be reported and discussed for this experiment, as analysis of the discriminability data (ANOVA) revealed nothing of significance other than performance impairments for inversion and incongruence. Correct reaction times had extreme values removed as in Experiment 1; 2.78% of correct trials were trimmed in this manner. Fig. 6 displays the raw results for reaction time, averaged across all subjects. Incongruent and congruent trials are indicated by black and white shapes, respectively. Upright triangles indicate upright trials, while upside-down triangles indicate inverted trials. Misaligned trials are denoted by squares. Fig. 6A considers only aligned trials and shows a decreased effect of incongruence on inverted faces. Fig. 6B considers only upright trials; except at −80 ms SOA, incongruence slowed response to misaligned faces much less than to aligned faces. For clarity, inverted misaligned trials are not shown; both incongruent and congruent trials were slower than corresponding trials that were only inverted or misaligned.

Data for each subject were normalized to zero mean and unit variance as in Experiment 1. These normalized values were used in a four-way ANOVA (SOA × alignment × orientation × congruence). With the exception of SOA × orientation ($F = 1.19$, $p > 0.3$), all main effects and two-way interactions were significant. SOA × alignment was only marginally significant ($F = 3.97$, $p < 0.05$), the others were significant at $p < 0.005$. Higher-order interactions were not significant except for alignment × orientation ×congruence ($F = 4.73$, $p < 0.05$).

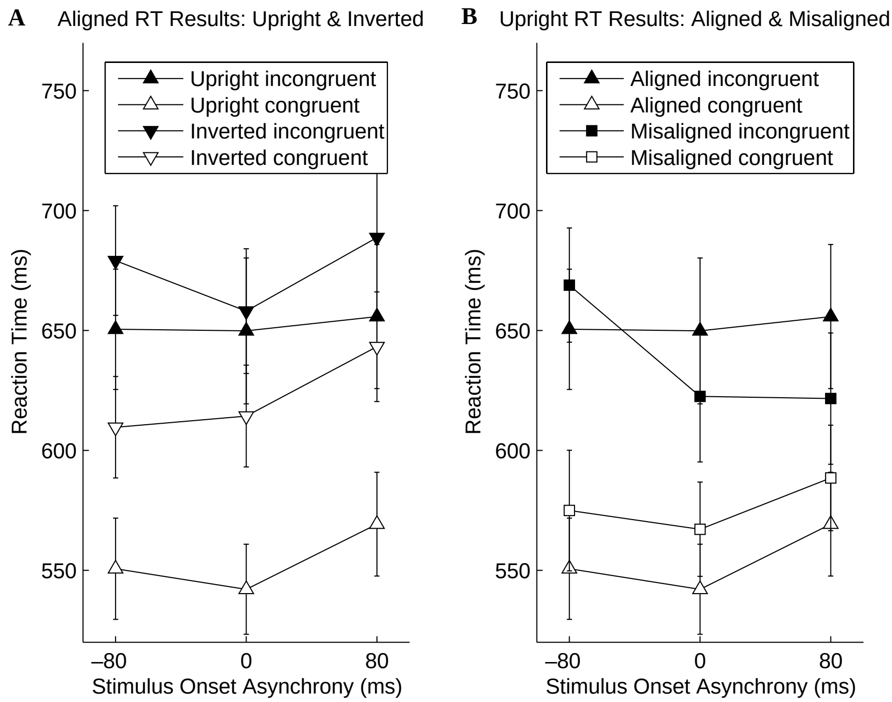

Fig. 6. Abscissa is SOA, ordinate is reaction time averaged across all subjects. Congruent trials are indicated by white shapes, incongruent trials by black shapes. (A) Upright (upwards-pointing triangles) and inverted (downwards-pointing triangles) trials for aligned faces. (B) Aligned (triangles) and misaligned (squares) trials for upright faces. Inversion and misalignment both slow response, as does incongruence. The effects are sub-additive, however, and misaligned incongruent faces are actually recognized more quickly than aligned incongruent faces.

The magnitude of the composite face effect was calculated as a difference of differences of reaction times, and we tested for effects of both misalignment and inversion on this measure. The results are shown in Fig. 7. Figs. 7A and B are basic measures of the composite face effect with respect to misalignment and to inversion, respectively. Figs. 7C and D each show the same measures as Figs. 7A and B, except that all trials considered in Fig. 7C were inverted and all trials considered inFig. 7D were misaligned. In other words, we examined the composite face effect with respect to one manipulation in faces whose configurations had already been disrupted by another manipulation. There were composite face effects at 0 and 80 ms SOA with respect both to misalignment and to inversion (Figs. 7A and B). There was, however, no effect with respect to a either manipulation when the other manipulation had already been applied (Figs. 7C and D)—for example, inverted faces did not show any composite face effect with respect to misalignment. At −80 ms SOA, there was a significant composite face effect with respect to inversion, regardless of the alignment of the halves. There was no effect with respect to misalignment at either orientation.

We omit from Experiment 2 the state-trace analysis used in Experiment 1 to demonstrate that the underlying psychological space is not one-dimensional. While the results from this experiment are consistent with a two-dimensional interpretation, there are only three data points per condition—too few from which to draw any meaningful conclusion.

3.3. Discussion

Using horizontally split faces, we reproduced the composite face effect for 0 and 80 ms SOA, using either inversion or misalignment to disrupt face processing. Either manipulation appears sufficient to disrupt the majority of processing that would otherwise bind an irrelevant face-half to a target that appeared simultaneously or previously. The interactions between alignment, orientation, and congruence revealed by the ANOVA reflect these findings: the effect of incongruence is reduced when faces are misaligned or inverted, but the effects of the two manipulations are not additive.

At −80 ms SOA there is no composite face effect with respect to misalignment, regardless of orientation. Why is there a composite face effect with respect to inversion at −80 ms SOA for both aligned and misaligned faces? One possibility is that an unattended upright bottom half can still activate some identity trace that persists over time and interferes with the required judgment. Our finding of a composite face effect with respect to inversion supports this: regardless of the alignment of the face parts, an incongruent upright bottom half interferes with identification of a top half appearing later. We would also expect to see no effect with respect to misalignment: the early irrelevant half primes the face-recognition system towards an incorrect judgment before the target even appears, whether it eventually shows up aligned or misaligned. Gauthier et al. (2003) have also demonstrated that there

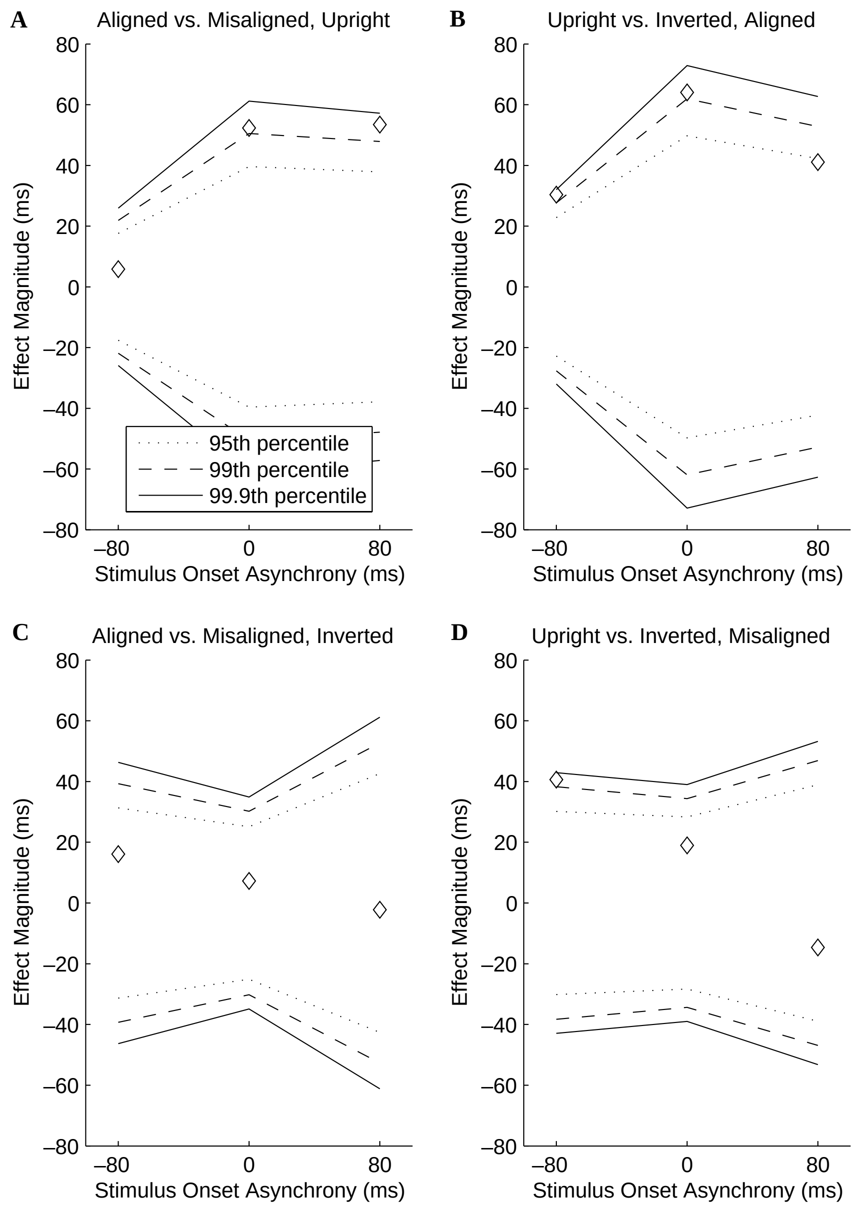

Fig. 7. (A and B) Show the basic reaction time composite face effect with respect to misalignment and inversion, respectively. Ordinate values are the magnitudes in ms, plotted at the three different SOAs. Dotted, dashed, and solid lines indicate, respectively, the 95th, 99th, and 99.9th percentiles given by a complete permutation test. (C and D) Show the same measures as (A) and (B), respectively, except that all trials considered in (C) were inverted and all trials considered in (D) were misaligned. There are significant effects at 0 and 80 ms in the single-manipulation cases, and at −80 ms with respect to inversion regardless of alignment.

can be interference from an irrelevant half of a horizontally split face presented in the study phase of a matching trial, even when that face-half is not even touching the target half. An incongruent face half can interfere with identification of its complement when it is presented first, despite spatial misalignment.

4. General discussion

Ingvalson and Wenger (2005) provided strong evidence that holistic and featural processing of faces proceed in parallel, with one or the other dominating depending on orientation, until some threshold is reached. Vinette, Gosselin, and Schyns (2004) recently suggested that different parts of a face are used effectively for identification at different latencies from stimulus onset. We propose that even the holistic interactions between the parts of a face admit new data as they arrive, rather than simply relating the parts that are present at a given moment. Holistic processing is initiated when an upright face or some sufficient part of a face is presented; properly configured evidence then accumulates in some “holistic processor,” even across asynchronies of 120 ms or more. Improperly configured features may be registered, but will not be inevitably bound to the face being processed.

That information from visual stimuli presented in rapid succession can interact perceptually is an old and well-established idea. Masking, priming, and the attentional blink (Raymond, Shapiro, & Arnell, 1992) are familiar examples. None of these, however, can account simply for our results. Any masking effect by the target on the distractor would reduce the distractor's influence. One would expect an upright target to be more salient and hence a more effective mask, however, the opposite pattern of interference was observed. For the same reason, these data are not well explained by the attentional blink phenomenon. Nor can the results at positive SOAs be attributed to priming by the distractor, as the distractor follows the target. We have shown that two parts of a face can interact across temporal gaps, and that the *nature of the interaction* depends on the properties of both parts.

Perrett, Oram, and Ashbridge (1998) proposed a more general accumulator model for object recognition at the cellular level, and described how it could account for observed effects of (among others) rotation, inversion, and misalignment. A similar accumulation of evidence model has recently been demonstrated at the single-neuron level for motion discrimination. Roitman and Shadlen (2002) describe the response of neurons in macaque cortical area LIP to random-dot motion. Spike rates of cells ramped up with continued exposure to motion in their preferred direction, until a response was made. Steeper ramping was correlated both with faster reaction times and stronger stimuli. Additional work (Huk & Shadlen, 2005) introduced a brief pulse of coherent motion to the background on some trials; these pulses, analogous to our irrelevant face halves, deflected the spike-rate ramp of a cell up or down depending on whether they agreed or disagreed with the cell's preferred direction.

Several studies have established that holistic processing is disrupted by inversion (Farah et al., 1998; McKone, Martini, & Nakayama, 2001). Recent evidence also suggests that holistic processing takes place automatically. Khurana, Smith, and Baker (2000) found that performance in a face-matching task was slowed when the faces had served as distractors in the previous trial, but that this effect vanished when the distractors were upside-down. Boutet et al. (2002) reported a composite face effect even when the faces in question were irrelevant to the task during which they were encoded (and in some cases actively interfered with it).

Under the proposed model, there is little holistic processing and hence little interference when the face parts are inverted. We should therefore see a composite face effect exactly when there is such interference in the corresponding upright cases. When a face interior is presented simultaneously with or preceding a face exterior, holistic processing is triggered, and the face exterior will (if incongruent) interfere with identification. A face top will also trigger holistic processing, but its identification will only be impaired by a simultaneous or following bottom if it is properly aligned. A face bottom will trigger holistic processing, biasing the system towards an identity and interfering with the identification of any incongruent face top that follows it.

Why might a face exterior have no impact on the identification of a following interior half, while a face bottom interferes with the identification of a top that appears soon after? Ellis, Shepherd, and Davies (1979) found an advantage for recognition of the inner parts of famous faces relative to the outer parts, and Moscovitch and Moscovitch (2000) describe an object agnosic who recognizes whole faces and face interiors as well as normal controls do, but fares much worse on face exteriors and inverted faces. Sergent (1984) examined interactions between different dimensions of photo-fit faces, and found that all significant interactions involved the configuration of the internal features. These results suggest that a face exterior in isolation is processed in a simple featural manner similar to common objects, and is therefore insufficient to initiate holistic face processing.

A leading exterior would not automatically activate holistic face processing; it might not activate any identity trace at all. Grill-Spector and Kanwisher (2005) report that basic-level classification happens concurrently with object detection but subordinate-level identification takes longer. The exterior face could be recognized as irrelevant and discarded from consideration before any identification takes place, and therefore not significantly affect the identification of the following interior. A face bottom, on the other hand, contains enough internal features that holistic processing, with its concomitant influence on the identification process, is inevitably set off.

Our results are in accord with a simple model positing that the holistic component of recognition is driven by an accumulation of properly configured evidence. The interactions between parts that are so important for face identification can survive asynchronous presentation. These interactions are evident even when parts are separated by 80 ms of noise, and so cannot be explained as ongoing processing of a holistic snapshot. Physiologically, these results suggest that the neural units responsible for identifying faces integrate information over a limited, but extended, time scale; a demonstration of such would be a strong complement to the present study. Meanwhile, we suggest that a complete understanding of the mechanisms underlying face processing requires consideration of both the spatial and temporal dimensions of the stimuli.

Acknowledgments

We thank James Tanaka and Fred Singer and members of the PEN network for their helpful comments and suggestions. J.M.S. was supported by the Burroughs-Wellcome Fund. D.L.S. was supported by the James S. McDonnell Foundation, NSF CRCNS 0423031, and NIH RO1-EY014681.

References

Boutet, I., Gentes-Hawn, A., & Chaudhuri, A. (2002). The influence of attention on holistic face encoding. *Cognition, 84*(3), 321–341.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*(4), 431–439.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482–498.

Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *Journal of the Anthropological Institute, 8*, 132–144.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Gauthier, I., Tanaka, J. W., & Brown, D. D. (2003). When misaligned faces are processed holistically [Abstract]. *Journal of Vision, 3*(9), 92a. doi:10.1167/3.9.92.

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*(2), 431–446.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152–160.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*(1), 65–74.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision-making. *Journal of Neuroscience, 25*, 10420–10436.

Ingvalson, E. M., & Wenger, M. J. (2005). A strong test of the dual-mode hypothesis. *Perception & Psychophysics, 67*(1), 14–35.

Khurana, B., Smith, W. C., & Baker, M. T. (2000). Not to be and then to be: Visual representation of ignored unfamiliar faces. *Journal of Experimental Psychology-Human Perception and Performance, 26*(1), 246–263.

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., & Abidi, M. A. (2005). Recent advances in visual and infrared face recognition: A review. *Computer Vision and Image Understanding, 97*(1), 103–135.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*(4), 259–272.

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review, 111*(4), 835–863.

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in roc space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin and Review, 3*(2), 164–170.

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

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology-Human Perception and Performance, 27*(3), 573–599.

Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology, 17*(1), 201–219.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition, 67*(1–2), 111–145.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*(5), 125–126.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology-Human Perception and Performance, 24*(4), 979–992.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22*(21), 9475–9489.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*(Pt. 2), 221–242.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science, 28*(2), 289–301.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 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*(2), 337–352.