



Different measures of holistic face processing tap into distinct but partially overlapping mechanisms

Isabelle Boutet¹ · Elizabeth A. Nelson¹ · Nicholas Watier² · Denis Cousineau¹ · Sébastien Béland³ · Charles A. Collin¹

Accepted: 3 June 2021 / Published online: 27 June 2021
© The Psychonomic Society, Inc. 2021

Abstract

Holistic processing, which includes the integration of facial features and analysis of their relations to one another, is a hallmark of what makes faces ‘special’. Various experimental paradigms purport to measure holistic processing but these have often produced inconsistent results. This has led researchers to question the nature and structure of the mechanism(s) underlying holistic processing. Using an individual differences approach, researchers have examined relations between various measures of holistic processing in an attempt to resolve these questions. In keeping with this, we examined relationships between four commonly used measures of holistic face processing in a large group of participants ($N = 223$): (1) The Face Inversion Effect, (2) the Part Whole Effect (PWE), (3) the Composite Face Effect, and (4) the Configural Featural Detection Task (CFDT). Several novel methodological and analytical elements were introduced, including the use of factor analysis and the inclusion of control conditions to confirm the face specificity of all of the effects measured. The four indexes of holistic processing derived from each measure loaded onto two factors, one encompassing the PWE and the CFDT, and one encompassing the CE. The 16 conditions tested across the four tasks loaded onto four factors, each factor corresponding to a different measure. These results, together with those of other studies, suggest that holistic processing is a multifaceted construct and that different measures tap into distinct but partially overlapping elements of it.

Keywords Face recognition · Holistic processing · Featural processing · Configural processing · Individual difference scores

The ability to recognize others is important for social interactions because it drives feelings of familiarity, approach behaviours,

Isabelle Boutet and Elizabeth A. Nelson contributed equally to this work.

Public significance statement Holistic processing, or the integration of facial features and their relations, has become a hallmark of what is considered ‘special’ about faces. We examined the nature and structure of the mechanism(s) underlying holistic face processing by measuring relationships between four commonly used measures of holistic processing. Results suggest that holistic processing is not a single process but a number of related ones, and that different measures of holistic processing tap into distinct subprocesses. This finding helps us better understand how our visual system interprets human faces, one of the most important categories of visual stimuli in our environment.

✉ Isabelle Boutet
iboutet@uottawa.ca

¹ School of Psychology, University of Ottawa, Ottawa, ON K1N6N5, Canada

² Department of Psychology, Brandon University, Brandon, MB, Canada

³ Faculté des Sciences de l'Éducation, Université de Montréal, Montréal, QC, Canada

and subsequent relationships. Following years of active debate, there is now general agreement that recognizing a person relies critically on the integration of facial features and their relations (reviewed by Maurer et al., 2002; Richler & Gauthier, 2014; Richler et al., 2012). This so-called holistic processing strategy is thought to be highly adaptive, arising from socially driven expertise in individuating face exemplars that are highly homogeneous (i.e., in the sense that all faces are made up of the same basic features positioned in the same configuration; e.g., Diamond & Carey, 1986; Leder & Bruce, 2000; McKone et al., 2007). Accordingly, there is evidence that faces are perceived more holistically than other objects, except in rare cases of expertise with nonface objects (Richler & Gauthier, 2014; but see McKone et al., 2007). Within this perspective, holistic processing is considered a hallmark of what makes faces ‘special’.

While holistic face processing has been the focus of hundreds of research articles, there is substantial heterogeneity in terminology, experimental paradigms, and underlying mechanisms associated with this processing strategy (Maurer et al., 2002; Richler et al., 2012; see Fig. 1). According to one model of holistic processing, face parts are integrated into a Gestalt or global template where the whole is greater than the sum of its

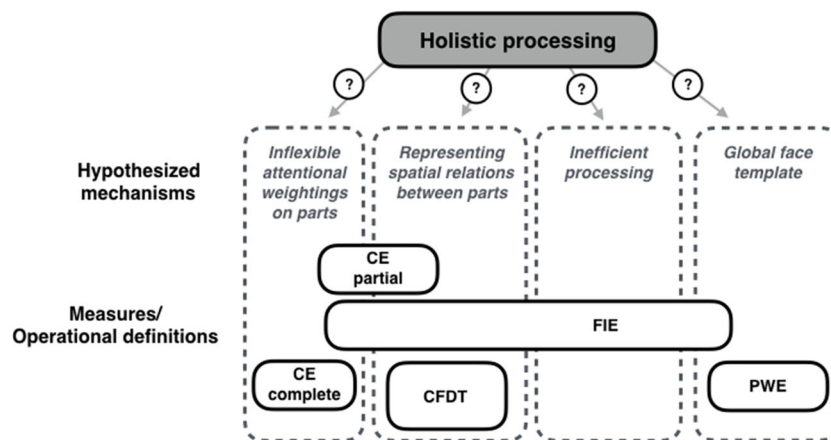


Fig. 1 Hypothesized mechanisms inspired from Richler et al. (2012). The dashed grey boxes represent mechanisms and the solid black boxes represent different measures of holistic processing. Different measures of holistic processing may tap in to a common mechanism, holistic processing. Different measures may tap into distinct mechanisms, as illustrated by the use of dashed grey boxes. Different measures may tap into partially

parts (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997). This mechanism is thought to be reflected in the Part-Whole Effect (PWE), whereby processing of face parts is superior when they are presented in whole faces versus in isolation (Tanaka & Farah, 1993; reviewed by Tanaka & Simonyi, 2016). According to the ‘configural’ model¹, faces are recognized on the basis of mechanisms that assess second-order relations between face parts (e.g., Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000). This mechanism is thought to be indexed by the Configural Featural Detection Task (CFDT; Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000). In the CFDT, participants have to detect differences in individual face parts or differences in the configuration between face parts. Whereas sensitivity to changes in configuration is superior in upright than inverted faces, sensitivity to changes in face parts is not (Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000). According to a third model, holistic processing reflects an inflexible or obligatory encoding strategy whereby all face parts are encoded together, which interferes with the ability to ignore irrelevant face parts (Richler & Gauthier, 2014). Support for this model comes from the Composite Effect (CE), where participants are asked to judge face composites made up of two half faces (top and bottom). When the two halves are aligned, participants have difficulty ignoring the bottom half, which interferes with their ability to judge the top half of a previously encoded face.

Whether these different mechanisms are distinct, partially overlapping, or manifestations of a common ‘holistic’

overlapping mechanisms, as illustrated when a measure box overlaps more than on mechanism. In the original illustration, the face inversion effect is positioned under sensitivity to configuration. We have illustrated it as an overlapping mechanism because the other measures in the figure are significantly reduced or even eliminated by inversion

processing mechanism remains unclear. Support for a single common underlying mechanism can be found in studies that show that different measures produce consistent findings across a given experimental manipulation or group difference. For example, performance is more greatly affected for faces than other objects for these four measures (e.g., Maurer et al., 2002; Richler et al., 2012). Another line of evidence comes from studies of prosopagnosia, a condition wherein individuals are exceptionally poor at recognizing faces (reviewed by Corrow et al., 2016). Individuals with prosopagnosia do not show the same pattern of results as healthy controls on most measures of holistic processing (reviewed by, e.g., DeGutis, Cohan, et al., 2013; Robotham & Starrfelt, 2018). More recently, researchers have borrowed from the psychometric tradition to examine the nature and structure of the mechanism(s) underlying holistic processing. In the face recognition literature, these studies are referred to as ‘individual differences’ studies to contrast with studies that focus on group-level effects. Individual differences studies take advantage of naturally occurring variations in performance to reveal patterns of association across tasks (see, e.g., Cooper et al., 2017; Wilmer, 2008). Confidence in the relationship between tasks is accrued across multiple studies over time (Strauss & Smith, 2009). As such, the individual differences approach requires replication of results from different data sets.

Our main goal was to contribute to this effort by examining relationships between four commonly used measures of holistic processing². By so doing, we are able to test the various models of the structure of holistic processing mechanisms.

¹ While the distinction between configural and holistic was made explicit in earlier studies (Maurer et al., 2002), the two concepts seem to have merged in more recent literature (Richler et al., 2012). This has led us, and others (e.g., Rezlescu et al., 2017), to incorporate these two mechanisms under the umbrella term ‘holistic’.

² We have addressed the question of whether holistic processing is linked to general face recognition abilities in a separate study (Nelson et al., 2016). The reader can turn to the following for more information: DeGutis et al. (2013b); Konar et al. (2010); McGugin et al. (2012); Richler, Cheung, and Gauthier (2011a); Richler et al. (2014, 2015); R. Wang et al. (2012).

This is because each model predicts different relationships between the outcomes of the different measures. In the following sections, we begin by introducing the four measures of holistic processing that were incorporated in this study, followed by a description of past studies that have examined the relationships, or lack thereof, between these different measures. We then provide details of the present study and the outcomes predicted by different models of the structure of holistic face processing mechanisms.

Measures and mechanisms of holistic face processing

The present study focuses on the following measures of holistic processing: the PWE, the CE, the CFDT, and the Face Inversion Effect (FIE). These measures were selected because they have been used in a large number of studies and have become standard manifestations of holistic face processing. All of these measures have been widely used in studies on development across the lifespan (e.g., Boutet & Faubert, 2006; Tanaka & Simonyi, 2016), in clinical populations (e.g., Lavallée et al., 2016; Watson, 2013), in individuals with prosopagnosia (e.g., DeGutis, Cohan, et al., 2013; Robotham & Starrfelt, 2018), and in comparative psychology (e.g., Brecht et al., 2017; Racca et al., 2010; Tomonaga, 2018; M.-Y. Wang & Takeuchi, 2017). All of these measures are reduced or eliminated for control nonface objects or inverted faces³. As such, all of them are believed to reflect what is ‘special’ about face processing.

The PWE (Tanaka & Farah, 1993; reviewed by Tanaka & Simonyi, 2016) refers to superior processing of face parts when they are presented in whole faces versus in isolation. The PWE suggests that face parts are integrated into a whole or Gestalt, and is therefore linked with a global template mechanism.

The CE (Young et al., 1987) refers to poor discrimination or recognition of the top half of a composite face made up of a top and a bottom from different individuals. The CE is diminished or eliminated when the two halves are misaligned. In terms of which mechanism(s) it indexes, the CE may reflect the formation of a global face template (Rossion, 2013; Young et al., 1987), a failure to selectively attend to face parts, and/or sensitivity to configuration (Richler et al., 2012; Richler, Cheung, & Gauthier, 2011).

Two paradigms have been developed to capture the CE, the so-called partial and complete designs, and there is still an active debate over which method best captures holistic

processing (Richler, Cheung, & Gauthier, 2011; Richler & Gauthier, 2014; Rossion, 2013). We used the partial design because historically it is the most widely used of the two designs, and as such, its inclusion will enable interpretation of our results across a broader range of existing literature (see Discussion for more details).

The CFDT (Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000) measures sensitivity to spatial relations between parts, which is consistent with a configural mechanism. In this paradigm, participants have to detect changes in the configuration between face features and changes in individual features. Detection of configural changes is affected by inversion, but not detection of changes in individual parts. While the ecological validity of the CFDT has been criticized (Taschereau-Dumouchel et al., 2010), we have included it because it is the only measure that taps directly into sensitivity to spatial relations, a concept that is central to the literature on holistic processing (e.g., Burton et al., 2015; Caharel et al., 2006; Maurer et al., 2002; Negrini et al., 2017; Rotshtein et al., 2007; Sergent, 1984; Yin, 1969), to face space theories (Valentine et al., 2016), and to computer-based recognition systems (Schwaninger et al., 2002). Moreover, brain correlates of face processing are sensitive to manipulations of spatial relations between face parts (e.g., Caharel et al., 2006; Negrini et al., 2017; Rotshtein et al., 2007). Finally, to our knowledge, no study has incorporated a measure that captures a configural mechanism when investigating the nature and structure of the mechanism(s) underlying holistic processing.

The FIE (Valentine, 1991; Yin, 1969) refers to the finding that discrimination and recognition of faces is disproportionately impaired by inversion as compared to nonface objects. Yin (1969) suggested that the FIE arises because faces are processed on the basis of relations between face parts. Many of the effects revealed by other measures of holistic processing are significantly reduced or even eliminated by inversion (Freire et al., 2000; McKone et al., 2013; Tanaka & Simonyi, 2016; but see, e.g., Susilo et al., 2013, for the CE). As such, the FIE might reflect any of the mechanisms associated with these measures (see Fig. 1). For example, the FIE might reflect a configural mechanism whereby it is difficult to extract second-order relations from inverted faces (Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000; Yin, 1969). Accordingly, the ability to detect configural changes is reduced by inversion in the CFDT.

Past research on the structure of the mechanism(s) underlying holistic processing

Three general hypotheses exist regarding the mechanisms that underlie measures of holistic processing. A commonly held assumption in this body of literature is that different measures

³ As previously mentioned, holistic processing can be elicited by objects of expertise. As such, the effects revealed by these paradigms can be found in rare cases with individuals who have developed expertise with individuation of other homogeneous nonface object categories via lifelong exposure or extensive training (see Richler & Gauthier, 2014; but see McKone et al., 2007).

are manifestations of a common ‘holistic’ processing mechanism (Rezlescu et al., 2017; Richler & Gauthier, 2014). Others have hypothesized that specific measures tap into distinct mechanisms, with some measures tapping into several mechanisms. For example, it may be argued that the FIE and the partial CE tap into overlapping mechanisms because both measures have been associated with sensitivity to configuration. A framework describing putative overlap between measures of holistic processing is shown in Fig. 1 (adapted from Richler et al., 2012). A third hypothesis is that different measures of holistic processing tap into distinct underlying mechanisms. This hypothesis is supported, by, for example, evidence that effects revealed by different measures emerge at different developmental time points (Maurer et al., 2002).

More recently, researchers have used an individual differences approach to compare different measures of holistic processing and shed light on these hypotheses. Their results are summarized in Table 1. The small number of studies conducted, together with inconsistencies in methodology across studies, makes it difficult to extract a unifying interpretation from these results. One difference relates to methods used to measure the CE—namely, using the partial CE (Rezlescu et al., 2017; R. Wang et al., 2012) versus the complete paradigm (DeGutis, Wilmer, et al., 2013). This research seems to suggest that the complete CE has more in common with other measures of holistic processing. However, the partial CE might overlap with measures that incorporate configural changes such as the CFDT (Richler et al., 2012), but this has not been investigated in any previous studies.

The studies cited in Table 1 also differ with respect to the methods used to index holistic processing. All of the paradigms included in these studies were originally intended to illustrate holistic processing at the group level (e.g., reviewed by Haxby et al., 2002; Maurer et al., 2002; McKone & Robbins, 2011). For example, in the FIE, holistic processing is illustrated by significant group-level differences in performance between upright faces and inverted faces. Two methods have been adopted to contrast conditions: subtraction scores (e.g., R. Wang et al., 2012; see also, e.g., Konar et al., 2010; Ross et al., 2015; Sunday et al., 2017) and residual or

regression scores, where scores are taken as residuals derived from regressing one condition against another (e.g., DeGutis, Wilmer, et al., 2013; Rezlescu et al., 2017; see Methods for more details). To date, there is no consensus on which method should be adopted in individual differences studies and differences in these methods may yield inconsistent results. It should also be noted that results derived from difference scores are difficult to interpret in individual differences studies (Hedge et al., 2017; Humphreys, 1990). For example, cognitive tasks such as the ones used herein often become well established because variability between subjects is low, which produces consistent results across studies. However, low between-subject variability produces low reliability for individual differences (Hedge et al., 2017). Given that it remains to be determined which method is best for individual differences studies, we computed both difference scores and residuals.

Interpretation of results reported in Table 1 is further complicated by the fact that conclusions are drawn on the basis of the statistical significance of correlations. Despite being statistically significant, some of these correlations do not account for much of the variability in the data. For example, the highly significant correlation reported by Rezlescu et al. (2017) between the FIE and the PWE only accounts for 7% of variance ($r = .28, p < .001$). This suggests that the relationship between the FIE and PWE is weak, which may explain discrepancies in findings across studies. Another issue with interpretation of correlations is that there is no satisfactory means of comparing correlation coefficients to determine if, for example, Measures A and B are more related to each other than Measures B and C. Moreover, computing multiple correlations across multiple measures requires adjustment of the significance level and in turn reduces statistical power. These issues highlight the need for an alternative analytical approach, such as factor analysis.

Finally, whether the procedures put in place to implement these different measures were adequate to capture holistic processing can be questioned. For instance, these studies did not systematically include a control condition to demonstrate the face-specificity of the holistic effects measured (McKone et al., 2013). This is crucial considering that these effects can be observed with nonface stimuli under certain circumstances (Meinhardt-Injac et al., 2016; Richler, Bukach, & Gauthier, 2009). For example, in the FIE, holistic face processing is illustrated by showing that inversion reduces performance for faces to a greater extent than for control objects (e.g., Valentine, 1991; Yin, 1969). It is this significant 2X2 interaction between stimulus category and image orientation that allows one to conclude that the effect is greater for faces than objects and hence reflects something ‘special’ about faces. Yet none of the aforementioned studies have included control conditions for all tasks being measured. To be fair, adding a control condition may seem unnecessary considering that control objects are unlikely to elicit the construct of interest, holistic

Table 1 Summary of findings on the relationship, or lack thereof, between measures of holistic processing

	PWE	CE partial	CE complete
CE partial	ns ^{a, b}		
CE complete	$r = .44, p < .01^c$		
FIE	$r = .28, p < .001^b$	ns ^b	

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect

^a R. Wang et al. (2012). ^b Rezlescu et al. (2017). ^c DeGutis, Wilmer, et al. (2013)

processing (Gauthier et al., 2018). Furthermore, the data collected on the control condition may be of limited utility for analysis of relationships between tasks because there is no easy solution for capturing group-level interactions at the individual level. Nonetheless, we feel that inclusion of a control condition is critical to establish the presence of the expected 2×2 interaction for each measure and hence to confirm that the procedures utilized are adequate to capture holistic *face* processing.

Present study

The main goal of the present study was to investigate the structure of the mechanism(s) underlying holistic processing. Some aspects of our study have been made similar to past research in order to facilitate cross-studies comparisons and to provide converging evidence, as is expected in the psychometric tradition (Strauss & Smith, 2009). We also introduced new elements to address some of the issues outlined above. First, we included a control in all of our measures to demonstrate that our procedures elicited the expected holistic effects. Including a control condition also allowed us to explore patterns of association that would otherwise have been missed. For example, if holistic processing of upright faces is mandatory (e.g., Boutet et al., 2002; Farah et al., 1998; Maurer et al., 2002; Sergent, 1984), then all conditions where full upright faces are shown should have something in common that is distinct from conditions where objects or inverted faces are shown.

Second, we adopted a factor analysis approach inspired by previous work on measures of executive functions (e.g., Miyake et al., 2000). Factor analysis is based on the notion that measurable and observable (i.e., manifest) variables can be reduced to fewer latent variables that share a common variance. As such, a factor analysis approach is ideal to provide evidence on the underlying structure of holistic face processing (Yong & Pearce, 2013). Submitting our data to an exploratory factor analysis allowed us to investigate whether holistic processing tap into a common underlying mechanism (H1), distinct mechanisms (H2), or partially overlapping mechanisms (H3). We adopted an exploratory factor analysis rather than confirmatory factor analysis to allow patterns of association to emerge from the data. This approach is well-suited for this area of research because there are many different models of how different face processing mechanisms might be related (e.g., Maurer et al., 2002; Richler et al., 2012). A one-factor solution would support the hypothesis that different measures of holistic processing tap into a common underlying mechanism (H1). A four-factor solution would support the hypothesis that different measures of holistic processing tap into distinct mechanisms (H2). Finally, any

solution where at least two measures load onto the same factor would support the notion that some measures overlap (H3).

Third, we not only focused on indexes of holistic processing derived from contrasting two conditions, but also explored relationships between the 16 conditions measured in this study (4 conditions \times 4 tasks). While this procedure is at odds with past research in that holistic processing is typically evidenced by contrasting two conditions, these analyses offer a means to sidestep problems associated with difference scores. Moreover, comparing the 16 conditions may reveal associations that offer a more suitable framework to interpret previous work in this field such as whether face parts and their spacing are represented by the same or different mechanisms (Yovel et al., 2014).

Finally, indexes of holistic processing were computed using both the subtraction and regression methods to examine whether the two methods yield consistent results and to compare their reliability and variability. Accumulation of evidence regarding the reliability and variability of measures is important to guide future individual differences studies because associations can be obscured by poor within-task reliability, and because these studies take advantage of variability in performance to reveal patterns of association (Cooper et al., 2017; Cronbach, 1957; Wilmer, 2008). We also ran analyses on accuracy, and reaction time to examine whether different dependent variables produce different patterns of results. For example, speed and accuracy may reflect distinct underlying mechanisms in face perception (Hildebrandt et al., 2010). In anticipation of our results, accuracy and reaction time variables produced consistent results. In the interest of space and to facilitate comparison with other studies, only the results of accuracy measures are reported in this article. Reaction time results are provided in the [Supplementary Materials](#) (Section 2).

While there are studies that have compared performance between two (DeGutis, Wilmer, et al., 2013; R. Wang et al., 2012) or three measures of holistic processing (Rezlescu et al., 2017), to the best of our knowledge, this is the first study to report results of factor analyses across four different paradigms, 16 conditions, and two different means of indexing holistic processing. It is also the first to include the CFDT in such analyses. As such, this study advances our understanding of the structure of holistic face processing mechanism(s), a key aspect of what makes faces ‘special’.

Method

The participant and materials information are identical for the four measures used in this study. As such, this information is provided first, followed by details regarding stimuli and procedure for each measure. Presentation order was randomized across participants, as was the order of conditions within each

measure. Participants took approximately 1 hour and 30 minutes to complete the four measures.

Participants

Undergraduate students from the University of Ottawa ($N = 223$) participated in this study, which was approved by the University of Ottawa Social Sciences and Humanities Research Ethics Board. Participants were awarded course credit for participation. The initial sample size was 240; however, due to the within-subjects design of this study, we deleted participants listwise who quit the study before completing all four tasks, or who did not correctly follow instructions.

In order to establish that the typical interaction effects were present in each of the four tasks, we ran four 2×2 repeated-measures analyses of variance (ANOVAs). We used G*Power software (Faul et al., 2007) to conduct a post hoc power analysis to determine statistical power ($1 - \beta$). We used the following input parameters: effect size $f = 0.1$, alpha of 0.05, total sample size of 223, one group, four measurements, a correlation among repeated measures of .56 (based on the average correlation coefficient as calculated in the correlational analyses), and nonsphericity correction of 1 because our data met the assumption of sphericity (see Results). Based on these parameters, statistical power is high, reaching .98.

Costello and Osborne (2005) outlined the best practices when conducting an exploratory factor analysis. With regard to sample size adequacy, they acknowledge that there are no ‘strict’ rules, and ‘more is better’. The rule of thumb most researchers use to determine a priori sample size for an exploratory factor analysis is a participant to item ratio of 10:1. Our participant to item ratio is 223:16 which equates roughly 14:1.

General methods

Across all measures, the face stimuli were male and the images were greyscaled. Face stimuli were taken from the Max Planck Institute for Biological Cybernetics database, and were scaled to 9×9 cm ($9^\circ \times 9^\circ$ at 57 cm). All faces were presented in the frontal (0°) view. Using Adobe Photoshop CS4 (Adobe Systems Inc., San Jose, CA; [Adobe.com](http://www.adobe.com)) and MATLAB 2010a (The MathWorks Inc., Natick, MA; <http://www.mathworks.com>), the background was removed and replaced with a uniform grey field. Ears and hair were removed. All stimuli were equated for mean luminance and root-mean-square (RMS) contrast. While some face stimuli were used in more than one task, there were never more than two repetitions across tasks. Stimuli used during practice were different than those used in the experimental conditions. Experiments were conducted on a series of identical Dell Optiplex 9010 computers with an i5 Intel core, running Windows 7 on 17-inch monitors with a screen resolution of $1,920 \times 1,080$ and refresh rate of 60 hertz.

For all measures, a two-alternative forced-choice (2AFC) procedure was used. Before being presented with a given measure, participants were given verbal instructions and shown examples of the stimuli, followed by practice trials. Before the experimental phase began, the instructions were presented again, at which point participants had the opportunity to ask for clarification.

Face Inversion Effect (FIE)

Stimuli

Eighty faces and 80 chairs were presented either upright or inverted (see Fig. 2). Chairs were used as a nonface category because they are typically seen upright, as faces are, and because they have been used before in the literature (e.g., Boutet et al., 2003; Stein et al., 2012; Yovel & Kanwisher, 2005). The authors created the chair images by taking pictures of chairs where at least two exemplars were of the same style. These similar exemplars were used as pairs, with one being either the target or the distractor.

Procedure

The task began with four practice trials (one for each condition). Participants were told verbally to respond as quickly as possible. This was done because this task is relatively easy and accuracy levels are subject to ceiling effects. A sequential matching paradigm was used with Stimulus Manipulation (face or object), and Orientation (upright or inverted) as variables. Four blocks of 40 trials were presented in random order as follows: 40 upright face trials, 40 inverted face trials, 40 upright object trials, 40 inverted object trials (total of 160 trials). The task took approximately 15 minutes to complete.

First, a single inspection image (face or object) was presented for 0.4 s. This was followed by an interstimulus interval of 0.1 s, during which a blank screen was presented. After this, two test images (two faces or two chairs, depending on the task condition) were presented. Images presented at inspection and test were either all upright or all inverted. The participants’ task was to identify which of the two simultaneously presented test images matched the inspection image by pressing a key. The test images remained on the screen until the participant responded. Accuracy and reaction time were recorded.

Part Whole Effect (PWE)

Stimuli

Methods (see Fig. 3) were inspired by Tanaka and Farah (1993). Stimuli were either whole faces or isolated face parts

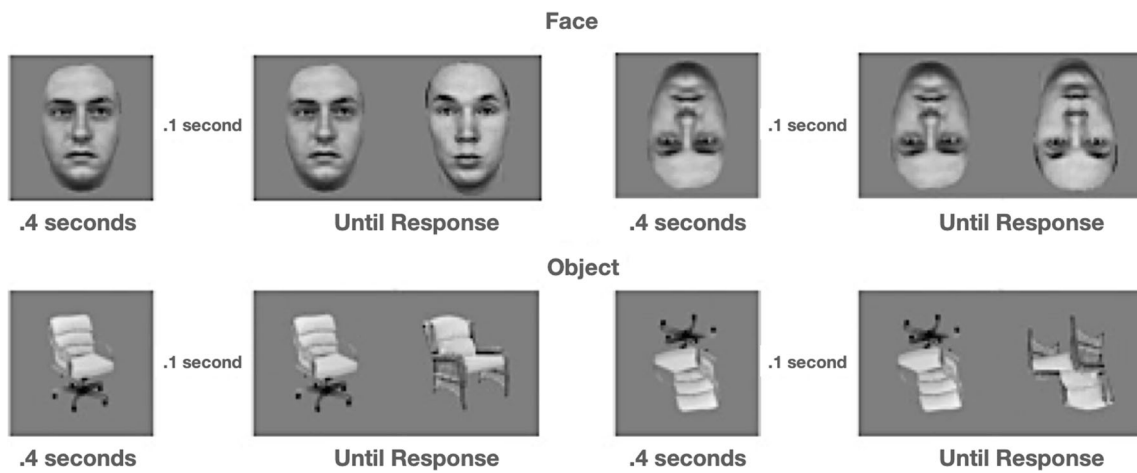


Fig. 2 Face Inversion Effect (FIE). The figure illustrates examples of stimuli, presentation times, and interstimulus intervals for upright and inverted trials

(features: eyes, nose, or mouth). To make the stimuli, 10 *base faces* were created. *Base faces* were created by replacing inner features (eye, nose, and mouth) by the same feature from different faces. The faces that were the source of the replacement features were different for each base face and not otherwise used in this experiment. This featural swap was done so that the base faces would not seem odder or modified than the modified faces that comprise the remainder of the stimuli in this task. From the 10 *base faces*, three modified versions were created by swapping either the eyes, or the nose, or the mouth, with that feature from another face. For example, Base Face A was used to create Version 1 with the eyes from Face B, Version 2 with the nose from Face C, and Version 3 with the mouth from Face D. Starting with 10 *base faces*, 30 modified faces were created (10 × 3 features). The *base faces* and three modified versions appeared in “Whole” conditions. The

facial features that were used to form the 10 *base faces* and the 30 modified faces appeared in the “Part” conditions.

Procedure

The task began with four practice trials (one for each condition). A sequential matching paradigm was used with Stimulus Manipulation (part vs. whole) and Orientation (upright vs. inverted) as variables. An equal number of “Part” and “Whole” trials were presented in random order. A total of 120 trials were presented in two blocks (60 trials per orientation). The task took approximately 10 minutes to complete.

The procedure for each trial was as follows. First, one base face was presented as the inspection face for 1.75 s followed by an interstimulus interval of 0.4 s, at which point two test faces (or test parts, both of the same feature) were presented

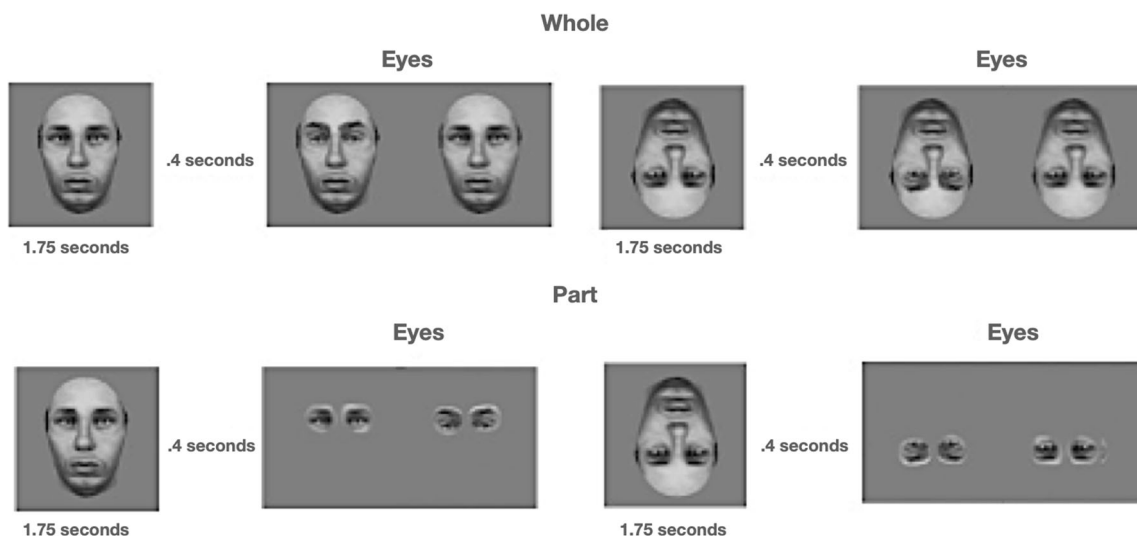


Fig. 3 Part Whole Effect (PWE). The figure depicts the trial structure for upright and inverted trials that pertain to eyes

simultaneously side-by-side. Above these two test faces (or parts), the word “Eyes” or “Nose” or “Mouth” appeared, cueing the participants to make their recognition judgment regarding that feature. This cueing is necessary in the “Whole” conditions so the participant knows which feature to compare between test and inspection faces. However, the cue also appeared in the “Part” conditions, despite it being obvious to which feature the participant must respond. Presenting the cue in both “Whole” and “Part” conditions equates for any distracting effects the cue might have. On the “Whole” trials, the participant’s task is to identify which of the two test faces contains the specified feature that matches the feature from the inspection face. On the “Part” trials, the participant’s task is to identify which of the two test parts (features) was contained within the inspection face. Participants responded by pressing a key. The test faces or features remained on the screen until the participant responded. Both accuracy and reaction time were recorded.

Composite Effect (CE)

Stimuli

Methods (see Fig. 4) were inspired by Young et al. (1987). Stimuli consisted of 240 composite face stimuli. Composite face stimuli were created by combining the top half of one face and the bottom half of a different face. Thus, all faces presented in this task were a hybrid of two faces. Composite faces were either presented with the two halves aligned, or with the top and bottom halves misaligned. In the case of the misaligned composite, the right edge of the bottom half of the face lines up with the middle of the nose from the top half of the face.

Procedure

The task began with instructions, followed by eight practice trials. Before the experimental phase began, the instructions were presented to the participant again, at which point he or she had the opportunity to ask for clarification. The researcher verbally emphasized to participants that they should selectively attend the top half of faces (middle of nose to forehead) in both the upright and inverted conditions.

A sequential matching paradigm was used with Stimulus Manipulation (aligned vs. misaligned) and Orientation (upright vs. inverted) as variables. An equal number of “Aligned” and “Misaligned” trials were presented in random order as follows: there were 120 ‘same’ trials and 120 different trials; these 120 trials contained 30 trials for aligned upright, 30 for aligned inverted, 30 for misaligned upright, and 30 for misaligned inverted (total of 240 trials). The task took approximately 20 minutes to complete.

Each trial began with one composite face (inspection face) presented for 0.5 s, followed by an interstimulus interval of 0.3 s, after which a second composite face (test face) was presented until response. Both faces were either upright or inverted, depending on the orientation condition, and were either aligned or misaligned, depending on the stimulus condition. The top halves of the inspection and test faces either matched or did not match, and the bottom halves never matched (i.e., were always different). Participants’ task was to judge if the top halves of the inspection and test faces (i.e., the part from the nose to the forehead, regardless of orientation) were the same or different. Participants indicated their response by pressing a key. Accuracy and reaction time were recorded. Performance on same trials only was used for the analyses.

Configural Featural Detection Task (CFDT)

Stimuli

Methods (see Fig. 5) were inspired by Freire et al. (2000). Face stimuli were either unmodified or modified. Modifications were either configural or featural. Configural modification involved moving a face part (eyes, nose, or mouth), either up or down 3 mm, which is enough to modify the spacing between parts without overly distorting the face. Featural modification means that a face part (eyes, nose, or mouth) was replaced with that same feature from a different face that was not used elsewhere in the experiment. Four modified base faces were created so that all face stimuli would look equally modified. These base faces were presented, along with the nine variations on each base face (i.e., both upward and downward configural manipulations for eyes, nose, and mouth, as well as the featural manipulation for each feature), making a total of 36 manipulated faces.

Procedure

There were eight practice trials (two for each condition). The task involved a matching paradigm with Stimulus Manipulation (configural vs. featural) and Orientation (upright vs. inverted) as variables. Two blocks of 48 trials were presented, one for the upright and one for the inverted condition (total of 96 trials). Within a block, half of the trials contained a featural manipulation and the other half contained a configural manipulation. For each type of manipulation, half of the trials presented two of the same face and half presented two different faces. Presentation order was randomized. The task took approximately 10 minutes to complete.

Each trial began with two faces presented simultaneously for 3.5 s. Participants indicated whether the two faces were the same or different using a key press. After 3.5 s, both faces were masked and participants were required to make a

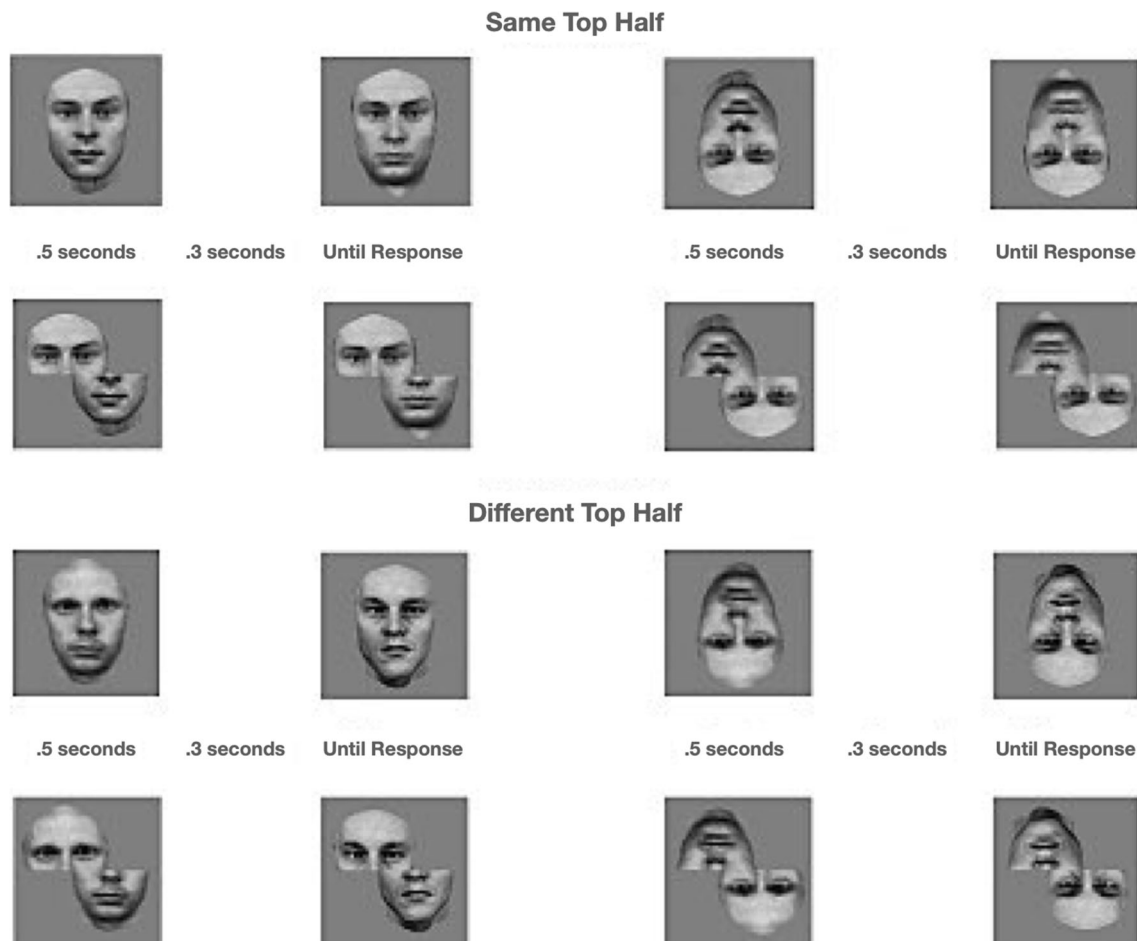


Fig. 4 Composite Effect (CE). The figure depicts the stimuli, presentation times, and the interstimulus interval for upright and inverted trials

judgment if they had not indicated their response yet. Accuracy and reaction time were recorded.

Dependent variables

Because the pattern of results was the same for accuracy and reaction time, we focus on accuracy to facilitate comparison with past research. Raw accuracy and reaction time data are available at <https://osf.io/69urz/> (<https://doi.org/10.17605/OSF.IO/69URZ>).

For each of our four tasks, we calculated an index of holistic processing by contrasting two key conditions using both the subtraction method and the regression method (DeGutis, Wilmer, et al., 2013). Calculating both types of individual difference scores allowed us to compare results obtained with each method and to report their reliability and variability. Table 2 lists the conditions that were contrasted for each task. Note that we chose to contrast conditions for which a difference in performance arises from a disruption in holistic processing, hence providing an index of holistic processing against a ‘baseline’. In the subtraction method, individual

performance on one condition is subtracted from performance in the other condition. In the residual method, performance in the condition of interest (e.g., the “Upright face” condition in the FIE) was regressed against performance in the baseline condition (e.g., “Inverted face” condition), and the regression line of best fit was obtained. The equation of the line of best fit was then used to calculate each participant’s expected score on the condition of interest, given their performance in the baseline condition. The values of the baseline condition were entered into the equation of the regression line of best fit to generate expected performance scores in the condition of interest. The expected scores were then subtracted from the actual performance scores on the condition of interest to obtain the residual value for each participant. Thus, residual difference scores indicate the degree to which a given participant’s performance is above or below the expected value. A high residual in, for example the FIE task, indicates that an individual is performing better on the “Upright face” condition trials than would be expected based on their performance on the “Inverted face” condition trials, which is consistent with a high degree of holistic processing.

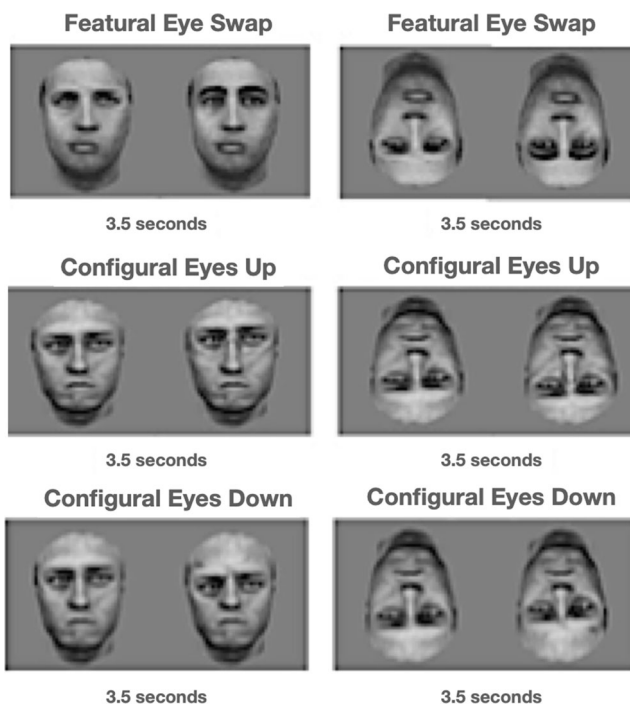


Fig. 5 Configural Featural Detection Task (CFDT). The figure depicts the stimuli, presentation times, and the interstimulus interval for upright and inverted trials that pertain to eye manipulation

Data cleaning

Data cleaning involved first calculating z -scores in all 16-task conditions to identify outliers. We chose the z -score cut-off value of ± 2.54 , which corresponds to an alpha of 0.01. Values 2.54 standard deviations above or below the mean score were considered to be outliers and were ‘tucked in’. This technique, which is similar to winsorizing (Hogg, 1979; Huber, 1981), entails replacing outlier values with the closest score that is within $z = \pm 2.54$ standard deviations of the mean. Less than 2% of all scores were tucked in.

We used Mahalanobis distances to identify the number of multivariate outliers in the accuracy data (eight participants

Table 2 Conditions contrasted to extract an index of holistic processing for each measure

Measure	Condition of interest	Baseline condition
FIE	Upright faces	Inverted faces
PWE	Upright whole	Upright part
CE	Upright aligned	Upright misaligned
CFDT	Upright configural	Inverted configural

FIE Face Inversion Effect, *PWE* Part Whole Effect, *CE* Composite Effect, *CFDT* Configural Featural Detection Task

out of 223, respectively). However, because we are interested in differences in performance across all conditions, because we believe these participants are valid examples demonstrating such differences, and because this small number of outliers is unlikely to offer an alternative explanation for our results, we did not remove participants who were multivariate outliers.

Data normality

We investigated skewness and kurtosis values in order to determine whether the accuracy scores between participants were normally distributed within each condition. Data were considered skewed if the estimated skew was more than twice the standard error of the skew, with the same logic applied for identifying kurtosis. In addition, we visually inspected the data using histograms with fitted normal curves and Q-Q plots. The results suggested that accuracy scores from all conditions within the FIE demonstrated negative (left) skew and were leptokurtic. Whole conditions within the PWE were negatively (left) skewed. All conditions within the CE were negatively (left) skewed, and the misaligned upright condition was also leptokurtic. Lastly, within the CFDT, the configural upright and featural upright task conditions demonstrated negative (left) skew and the configural inverted condition was positively (right) skewed. In light of recent findings that log transformation has negative consequences on subsequent data analyses (Feng et al., 2014), and in order to facilitate interpretation of our results, we elected to leave our data untransformed and use non-parametric correlation tests. Moreover, various violations of normality across measures would have required different transformations, making interpretation very challenging.

We assessed multivariate normality by plotting the Mahalanobis distance values against estimated chi-square quantiles in a Q-Q plot. Based on visual inspection and given that we elected to leave in the 8 multivariate outliers, we determined that accuracy scores did not meet the assumption of multivariate normality.

Results

Analysis of group effect interactions

In order to ascertain that the expected interaction effects for each measure were replicated, we conducted a series of 2×2 repeated-measures ANOVAs for each of the four measures. The results are consistent with the typical findings for each measure. These results point to the adequacy of our procedures for capturing holistic processing as defined by past research for all four measures. Descriptive statistics and

ANOVA tables are provided in the [Supplementary Materials](#) (Section 1: Group Effect Interactions).

Do the four measures of holistic processing tap into a common mechanism (H1), distinct mechanisms (H2), or partially overlapping mechanisms (H3)?

Exploratory factor analysis (EFA) was used to investigate the three hypotheses outlined in the introduction. EFA is used to reveal factors that emerge from variation in the data and the relationship between them, while maintaining conservative factor retention parameters. First, we determined that our data showed multicollinearity by checking the variance inflation factor in an iterative fashion such that performance on each condition was regressed against performance on the remaining 15 conditions using linear regression. Multicollinearity was expected because performance is correlated across conditions within each measure, as it should be if they are measuring the same underlying factor.

We used a direct Oblimin rotation (Osborne & Costello, 2009; Yong & Pearce, 2013) because correlation analyses revealed moderate to high correlations within each task, and weak correlations between tasks. In accordance with Tabachnick and Fidell (2001), we suppressed factor loadings below 0.32. The tables with all factor loadings are provided in [Supplementary Materials](#) (Section 1: Complete Factor Loadings). Results for the subtraction method are not presented because the Kaiser–Meyer–Olkin Measure of Sampling Adequacy was low (.51) and Bartlett’s test of sphericity was not significant, $\chi^2(6) = 8.86, p = .182$. The correlation matrix for subtraction-based scores is shown in Table 3.

The correlation matrix and EFA factor loadings for indexes of holistic processing derived from regression scores are shown in Tables 4 and 5. The Kaiser–Meyer–Olkin Measure of Sampling Adequacy was below the suggested cut off value of .50 (.44) so we investigated the anti-image correlation matrix. The diagonal values of the anti-image correlation matrix were approaching .50, suggesting there was likely adequate collinearity among the variables. Bartlett’s test of sphericity was significant, $\chi^2(6) = 20.21, p < .01$. Regression scores

Table 3 Correlation matrix for indexes of holistic processing (subtraction scores)

	FIE	PWE	CE	CFDT
FIE				
PWE	.00			
CE	-.03	.00		
CFDT	.07	.12	-.15*	

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task. * Significant at $p < .05$

Table 4 Correlation matrix for indexes of holistic processing (regression scores)

	FIE	PWE	CE	CFDT
FIE				
PWE	-.02			
CE	-.05	.13		
CFDT	.05	.18**	-.05	

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task. ** Significant at $p < .01$

loaded onto two factors, with PWE and CFDT loading together; this factor accounts for 13.5% in variance. CE loaded onto the second factor, which accounted for 6% of variance. These results suggest that PWE and CFDT measure an overlapping construct that is distinct from CE. Results with regards to the FIE are more difficult to interpret because of ceiling effects in two conditions for this measure.

Relationships between 16 individual conditions

Table 6 illustrates the correlation matrix for the 16 conditions tested across the four measures. As can be seen from this table, correlations within measures are generally larger than correlations between measures. Table 7 illustrates results of the EFA for the 16 conditions. Each task loads onto a separate factor. The PWE (Factor 1) accounted for most of the variability in the model (30%), followed by the FIE (Factor 2, 16%). The four factors accounted for 61% of the total variance explained.

Reliability and variability of the measures

Internal reliability was measured using Guttman’s λ_2 . We used this measure because it has previously been employed in individual differences work on face processing (Sunday et al., 2017) and therefore facilitates comparisons across studies. Guttman’s λ_2 is more robust than Cronbach’s α in instances where the measures are comprised of multiple factors (Callender & Osburn, 1979). Guttman’s λ_2 was calculated for

Table 5 EFA for indexes of holistic processing (regression scores)

Measure	Factor Communalities			
	1	2	Initial	Extraction
FIE			.00	.01
PWE	.46		.07	.27
CE		.46	.01	.20
CFDT	.55		.07	.29

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task

Table 6 Correlation matrix for the 16 conditions

FIE	Face Upright		.53	.41	.35	.22	.15	.12	.11	.18	.16	.29	.19	.26	.14	.20	.18
	Face Inverted	.53		.39	.35	.31	.30	.20	.13	.24	.13	.32	.20	.24	.22	.33	.25
	Object Upright	.41	.39		.34	.15	.10	.11	.06	.18	.09	.20	.14	.02	.09	.06	.17
	Object Inverted	.35	.35	.34		.12	.07	.14	.12	.13	.08	.15	.11	.08	.01	.14	.14
PWE	Whole Upright	.22	.31	.15	.12		.44	.44	.33	.24	.12	.25	.15	.26	.20	.30	.22
	Whole Inverted	.15	.30	.10	.07	.44		.42	.47	.19	.07	.17	.11	.22	.29	.36	.35
	Part Upright	.12	.20	.11	.14	.44	.42		.50	.17	.14	.13	.12	.22	.24	.24	.23
	Part Inverted	.11	.13	.06	.12	.33	.47	.50		.16	.04	.14	.06	.16	.26	.25	.25
CE	Aligned Upright	.18	.24	.18	.13	.24	.19	.17	.16		.55	.59	.50	.10	.17	.11	.16
	Aligned Inverted	.16	.13	.09	.08	.12	.07	.14	.04	.55		.55	.73	.10	.09	.18	.15
	Misaligned Upright	.29	.32	.20	.15	.25	.17	.13	.14	.59	.55		.57	.17	.12	.22	.22
	Misaligned Inverted	.19	.20	.14	.11	.15	.11	.12	.06	.50	.73	.57		.13	.08	.21	.18
CFDT	Configural Upright	.26	.24	.02	.08	.26	.22	.22	.16	.10	.10	.17	.13		.42	.46	.35
	Configural Inverted	.14	.22	.09	.01	.20	.29	.24	.26	.17	.09	.12	.08	.42		.49	.43
	Featural Upright	.20	.33	.06	.14	.30	.36	.24	.25	.11	.18	.22	.21	.46	.49		.59
	Featural Inverted	.18	.25	.17	.14	.22	.35	.23	.25	.16	.15	.22	.18	.35	.43	.59	
	Face Upright	Face Inverted	Object Upright	Object Inverted	Whole Upright	Whole Inverted	Part Upright	Part Inverted	Aligned Upright	Aligned Inverted	Misaligned Upright	Misaligned Inverted	Configural Upright	Configural Inverted	Featural Upright	Featural Inverted	
	FIE				PWE				CE				CFDT				

Note: Correlations of $r = .156$ and above were significant at the $p < .01$ level. Correlations of $r = .111$ and above were significant at the $p < .05$ level. $N = 223$. FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task

indexes of holistic processing using the steps outlined by DeGutis, Wilmer, et al. (2013). Results are shown in Table 8 for indexes of holistic processing and Table 9 for the 16

Table 7 EFA for the 16 conditions

Measure	Factor Communalities					
	1	2	3	4	Initial	Extraction
FIE	Face Up				.91	.93
	Face Inv				.81	.82
	Object Up				.91	.93
	Object Inv				.86	.88
PWE	Whole Up	.53			.40	.41
	Whole Inv	.65			.41	.48
	Part Up	.70			.41	.49
	Part Inv	.76			.41	.53
CE	Aligned Up		.66		.47	.48
	Aligned Inv		.86		.61	.70
	Misaligned Up		.66		.51	.55
	Misaligned Inv		.85		.61	.70
CFDT	Configural Up			-.62	.39	.39
	Configural Inv			-.57	.37	.38
	Featural Up			-.88	.53	.71
	Featural Inv			-.66	.42	.47

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task, Up Upright, Inv Inverted

conditions. Standard deviations are also reported as an index of variability. Internal consistency reliabilities of difference scores were lower than the reliabilities for each condition. Internal consistency reliability was higher for the regression than the subtraction method. Variability was comparable across the 16 individual task conditions and across the indexes of holistic processing derived from the PWE, CE, and CFDT tasks. Lower variability in the FIE task may arise from near ceiling effects in this condition (see Supplement: Group Effect Interactions).

Table 8 Reliabilities and variability (standard deviation [SD]) of the four indexes of holistic processing derived from each task for the subtraction and regression methods

Measure	Subtraction		regression	
	Guttman's λ_2	SD	Guttman's λ_2	SD
FIE	.85	.05	.88	.05
PWE	.27	.12	.41	.13
CE	.59	.14	.66	.14
CFDT	.20	.14	.48	.12

FIE Face Inversion Effect, PWE Part Whole Effect, CE Composite Effect, CFDT Configural Featural Detection Task

Table 9 Reliability and variability (standard deviation [*SD*]) of the 16 conditions

Measure	Condition	Guttman's λ_2	<i>SD</i> Subtraction
FIE	Upright Faces	.93	.09
	Inverted Faces	.93	.10
	Upright Objects	.94	.10
	Inverted Objects	.93	.09
PWE	Upright Part	.58	.11
	Upright Whole	.61	.11
	Inverted Part	.51	.11
	Inverted Whole	.58	.12
CE	Upright Misaligned	.82	.15
	Upright Aligned	.84	.16
	Inverted Misaligned	.78	.15
	Inverted Aligned	.79	.15
CFDT	Upright Configural Cong	.67	.13
	Upright Featural	.63	.10
	Inverted Configural	.40	.12
	Inverted Featural	.62	.12

FIE Face Inversion Effect, *PWE* Part Whole Effect, *CE* Composite Effect, *CFDT* Configural Featural Detection Task

Discussion

Over the past 30 years, the idea that faces elicit holistic processing has become a central tenet of the face recognition literature. Yet there is substantial heterogeneity in the terminology, paradigms, and mechanisms associated with this processing strategy (Richler et al., 2012). The present study examined relationships between four commonly used measures of holistic processing. This allowed us to test various models of the structure of holistic processing mechanisms. Relationships between measures were examined for the first time using a factor analysis approach. This approach addresses limitations inherent to correlations between pairs of measures (Yong & Pearce, 2013). It also offers a better tool for testing hypotheses drawn from the literature, such as potential overlap between certain measures (see Fig. 1). In keeping with the psychometric tradition, we focused on widely used procedures that capture key aspects of holistic processing so that our data would have the widest possible applicability with respect to accumulation of evidence over time (Strauss & Smith, 2009). We also provide information on the reliability and variability of each measure, which will be useful for researchers who wish to utilize them in the future.

The EFA solution for indexes of holistic processing provided a two-factor solution, with the PWE and CFDT loading into one factor and the CE loading into another factor. These results contradict the notion of a common holistic mechanism (H1). More specifically, they suggest that the PWE and CFDT

tap into an overlapping construct that is distinct from CE (H3). The model accounted for a modest portion of variance in the data (20%), which is consistent with past research where correlation coefficients between measures of holistic processing is typically low. It may come as a surprise that the PWE and CFDT loaded onto the same factor considering that these two measures capture seemingly different concepts, with the PWE reflecting analysis of faces as global templates and the CFDT reflecting sensitivity to spatial relations. Historically, these two concepts have been central to the literature on holistic processing since its very beginning (e.g., Maurer et al., 2002; Sergent, 1984; Valentine et al., 2016; Yin, 1969) and their loading onto the same factor suggests that they reflect overlapping processing mechanisms. One possibility is that this factor reflects improved processing of parts, including their relations, in the context of a whole face (Sunday et al., 2017). In the PWE, the presence of a whole face may introduce information about a given part's relation to other parts and in turn facilitate its processing. In the CFDT, judging a parts' relation to other parts (their spatial arrangement) may be more efficient in the upright than the inverted condition because relational information is more efficiently extracted from upright than inverted faces (e.g., Richler, Mack, et al., 2011; Yin, 1969). Alternatively, the PWE and CFDT factor may not reflect holistic processing but instead similarity in task demands in these two tasks. For both measures, participants were asked to judge individual face parts. However, this interpretation is challenged by the finding that individual conditions where parts were judged did not load together in the EFA with 16 conditions.

Consistent with our results, the PWE was not significantly correlated with the CE (partial design) in Rezesescu et al. (2017) and R. Wang et al. (2012). In contrast, performance on the PWE was significantly correlated with the CE (complete design) and the FIE in DeGutis, Wilmer, et al. (2013). Discrepancies for the CE across studies likely arise from the distinction between the complete and partial designs. Briefly, the partial design measures the difference in performance between *same aligned* and *same misaligned* trials where the top half of the composite is the same as a previously presented face. The complete design measures a congruency effect as the difference in performance between congruent trials, where the correct response for the top half of the composite is the same as that for the bottom half, and incongruent trials, where providing the correct response for the top half of the composite requires ignoring the response for the bottom half. It has been argued that the partial design taps into the formation of a global face template whereas the complete design taps into a failure to selectively attend to face parts (e.g., Richler & Gauthier, 2014; Rossion, 2013; Young et al., 1987). This might explain why the PWE relates to the complete CE but not the partial CE. Also worth noting is the finding of a significant correlation between the partial CE and CFDT task in

our study when holistic processing was measured using difference scores. This finding, which must be viewed with caution given the poor reliability of difference scores, is consistent with the notion that the partial CE taps into a configural mechanism (Richler et al., 2012). In any case, it appears that these two designs measure different constructs (Richler & Gauthier, 2014) with the complete CE being more closely related to the PWE than the partial CE. Additional evidence directly comparing the two versions of the CE to each other as well as other measures is needed to corroborate this interpretation.

With respect to the FIE, we were surprised that it did not load onto a common factor with the CFDT considering that for both measures, the index of holistic processing was computed by contrasting upright and inverted conditions. Results for the FIE should be interpreted with caution considering that performance was near ceiling. With respect to past research on the FIE, results have been inconsistent. In DeGutis, Wilmer, et al. (2013), the FIE correlated with the PWE. In Rezlescu et al. (2017), it did not correlate with any other measure. Interpretation of results pertaining to the FIE is further complicated by the debate over whether inversion has a qualitative or quantitative impact on face processing (e.g., Meinhardt et al., 2019; Rakover, 2013; Richler, Mack, et al., 2011; Rossion & Boremanse, 2008). Future research on the FIE is therefore needed to improve our understanding of the mechanisms that underlie this effect and its relationship, or lack thereof, with other measures of holistic processing, especially since reliability was highest for the FIE.

The EFA model for the 16 conditions yielded four factors with the four conditions within each task loading onto a separate factor. The correlation matrix revealed stronger correlations between conditions *within* each task than between theoretically equivalent conditions *across* tasks. These results again contradict the notion of a common holistic mechanism (H1). They support the existence of distinct holistic mechanisms (H3) with different measures tapping into a distinct facet of holistic processing. One possibility is that task instructions, together with exposure to a few trials results in participants adopting a cognitive set that carries over from one condition to the next within a given task. Similar effects have been reported elsewhere (Boutet et al., 2017; Richler et al., 2009b). For example, Richler et al. (2009b) have shown that presentation of whole faces prior to presentation of nonface objects can induce holistic processing of these objects. These carry-over effects were influenced by the procedures employed with blocked conditions being less likely to elicit carry-over effects than randomized presentations (see also Boutet et al., 2017). Whether blocking versus random order had a similar effect on the results of the present study is unclear because procedures varied across measures (trials were blocked by orientation in the CE and PWE, by condition in the FIE, and completely randomized in the CFTD task). In any case, these carry-over

effect do not appear to be due to the particular procedures used in this study since high within task correlations have been also been reported for the CE (Gauthier et al., 2018) and PWE (Sunday et al., 2017).

It may be argued that reporting relationships between individual task conditions is not informative because holistic processing is typically illustrated by interaction effects that have been operationalized as difference scores in individual differences research. To the contrary, we feel that reporting these high within-task correlations is critical because it suggests that each task triggers a unique processing mechanism (H2) that carries over from one task condition to the next. We attempted to devise a mathematical method to extract variability due to carry-over effects, but discontinued this effort due to problems accounting for interactions between conditions (an *in progress* formula is available upon request). Future research should report correlations between task conditions to reveal possible carry-over effects that can obscure relationships between different measures of holistic processing. These carry-over effects seriously limit the potential to find associations between different tasks and offer a plausible explanation for the finding that theoretically similar conditions are not correlated. For example, one might have expected the different conditions where the task at hand requires processing of individual features to be related to each other, but this was not the case. Moreover, carry-over effects limit our ability to examine predictions arising from the literature such as whether face parts and their spacing are represented by different mechanisms (Yovel et al., 2014) and whether all conditions with upright faces trigger mandatory holistic processing (e.g., Boutet et al., 2002; Farah et al., 1995; Maurer et al., 2002; Sergent, 1984). Finally, high within-task condition correlations may explain low correlations between measures of holistic processing (DeGutis, Wilmer, et al., 2013; Rezlescu et al., 2017; R. Wang et al., 2012) and between measures of holistic processing and face recognition ability (e.g., Konar et al., 2010; Rezlescu et al., 2017; Richler et al., 2014).

In the present study, we devoted much effort to design measures with comparable procedures. All tasks involved simultaneous or short-term sequential matching procedures, and all of the tasks required participants to perform a same/different discrimination task, making it unlikely that differences in these aspects of the task could explain our results. Another factor that can influence relationships between measures of face processing is number of faces used across trials. Richler and colleagues have argued that stimulus repetition can contaminate measures of holistic processing by inducing a part-based strategy whereby participants “focus attention on features that are particularly diagnostic to distinguish among a set of repeating face *parts*” (Richler et al., 2015, p. 2), albeit how many repetitions are required to engage this part-based strategy has yet to be specified. In the context of our study, there were few repetitions (e.g., 4 repetitions of face halves in

the CE) within tasks and none across tasks. It is therefore unlikely that stimulus repetition accounted for the results of our study. This explanation is also difficult to reconcile with the EFA solution since one might expect a part-based strategy induced by stimulus repetition to generalize across measures. Still, it is advised that as many different faces as possible be included in future studies to avoid this confound. Finally, we think it unlikely that spurious results were obtained because of the presence of outliers considering that one of the strengths of the present study is the large number of participants tested (223) and the small number of outliers.

Reliability and variability of the measures

Individual differences research relies on individual variability to extract shared variance among measures. Moreover, reliability of measures is important because low reliability can obscure relationships between measures. Hence, choosing measures for future individual differences research should be driven in part by information on their reliability and variability. Overall, our versions of the tasks produced good to acceptable reliability scores with Guttman's lambda values ranging between .40 to .94 for performance on the 16 task conditions and between .41 to .88 for regression scores. These reliability scores are equivalent or superior to those reported in other studies (e.g., DeGutis, Wilmer, et al., 2013; Rezlescu et al., 2017; Sunday et al., 2017; R. Wang et al., 2012), probably because of the larger number of trials per condition in the present study (Sunday et al., 2017). Reliability of subtraction scores was poorer than for regression scores, supporting the use of the regression for indexing holistic processing at the individual level (DeGutis, Wilmer, et al., 2013; but see Ross et al., 2015).

Looking at the data as a whole, it is difficult to suggest one or two measures for future individual differences research. For example, the FIE produced the highest internal reliability but had low variability because of ceiling effects and hence cannot be recommended in its current version. The CE had good reliability and variability, but it did not account for much of the variance in either the four indexes (6%) or the 16 conditions solutions (10%). Drawing inferences across studies is also difficult because the procedures used vary widely. Efforts at developing standardized versions of these tasks for individual differences work are currently underway (Richler et al., 2014, 2015). However, additional research across laboratories is needed to accumulate evidence on the validity and reliability of these measures. It might be useful for researchers to run pilot studies to verify that their measures have acceptable reliability prior to collecting data on relationships between measures. We report informally that reliability indexes were higher for efficiency scores that combined accuracy and reaction time than for accuracy scores. This suggests that future work on individual differences in holistic processing

should consider combining accuracy and speed in their analyses (see, e.g., Rate Correct Score; Gauthier et al., 2018; Vandierendonck, 2017), especially when accuracy scores are limited by ceiling effects. Combining accuracy and RT data may be more valid considering that face processing involves a speed dimension (Hildebrandt et al., 2010; Wilhelm et al., 2010).

Some studies have adjusted correlations between tasks by measuring a theoretical correlation that takes into account the reliability of each measure (DeGutis, Wilmer, et al., 2013; Rezlescu et al., 2017). While we recognize that unreliable measures cannot be expected to correlate strongly with one another, whether and how correlations should be disattenuated is still debated (see, e.g., Winne & Belfry, 1982). Issues associated with disattenuated correlations in the present study include overcorrected correlations (correlations greater than 1) that are difficult to interpret as well as problems with comparing disattenuated correlations with noncorrected EFA solutions. As a result, we do not report disattenuated correlations in the body of this article⁴.

Implications

What emerges from ours and others' findings (e.g., DeGutis, Wilmer, et al., 2013; Rezlescu et al., 2017; R. Wang et al., 2012) is that holistic processing is not a broad overarching mechanism that gives rise to the group-effect interactions elicited by these different measures. Instead, our results provide support for the hypothesis of overlap between distinct holistic mechanisms (H2, H3) (the PWE and CFDT in the present study, the PWE and complete CE in DeGutis, Wilmer, et al., 2013). While some measures can overlap, commonality among measures may not always be driven by holistic processing per se, but instead by other cognitive processes depending on task demands. As such, different measures should not be used interchangeably. Developmental changes in these other cognitive domains may explain why some of these effects produce inconsistent developmental trajectories across task (Boutet & Faubert, 2006; Maurer et al., 2002; Meinhardt-Injac et al., 2016).

One question that arises is how these measures can tap into a distinct mechanism that carried over from one condition to another and yet yielded significant interactions at the group level. The answer to this question may come from research with clinical or developmental populations. For example, older adults show poor face recognition ability alongside preserved holistic processing, suggesting that holistic processing is necessary, but not sufficient for correct identification of faces (Boutet & Meinhardt-Injac, 2018; Boutet et al., 2020; see also Watson, 2013, for autism and schizophrenia). These

⁴ The amount of shared variance between tasks remained modest even after disattenuation (see Supplement: Disattenuated correlations).

findings have been interpreted as evidence that face recognition proceeds in a ‘global to local’ fashion with holistic processing taking place in face selective areas, followed by analysis of local features via reentrant connections to lower-level visual areas where small receptive fields are well suited for processing finer details that distinguish individual exemplars (e.g., Edelman & Gally, 2013; Jacques & Rossion, 2009; Musel et al., 2012; Peters et al., 2018; Sugase-Miyamoto et al., 2011). It might therefore be argued that while the first ‘global’ or holistic processing stage is elicited by all of these measures, individual differences in performance reflect the second ‘local’ stage. If this were true, then individual differences in measures of holistic processing may instead reflect variations in the ability to process the featural aspects of faces (see also Jaworska et al., 2019). This could explain the lack of relationship between measures found here and elsewhere. It would be interesting for future research to devise means of examining this hypothesis by comparing individual performance on measures of ‘featural’ processing. Whether relationships between measures arise from holistic processing per se or from similarities in task-demands should also be considered.

Limitations

First, our results may be limited by the procedures used to operationalize the four measures of holistic processing. While we have implemented versions of each measure that were similar to the original authors’ implementations, some changes were made to make the tasks more comparable. Second, there is much debate over which version of the CE is most valid (Richler & Gauthier, 2014; Rossion, 2013). On one hand, our results support the utility of the partial CE: the paradigm produced the expected group-level interaction effect and the correlation between the partial CE and the CFDT was statistically significant for subtraction scores, which supports the claim that the partial CE taps into a configural mechanism (Richler et al., 2012). On the other hand, the low reliability of subtraction scores limits interpretation of this finding. Moreover, performance on the CE was relatively low in our study, and performance in the aligned upright and aligned inverted conditions was almost identical. This contrasts with typical results for the partial CE (e.g., Robbins & McKone, 2003; Wang et al., 2019) and may have decreased the magnitude of holistic processing triggered by the task (Richler et al., 2009a; Wang et al., 2019). Additional research comparing the partial and complete CE with each other, and with other measures of holistic processing, is therefore needed to resolve this debate.

Third, one might question the conditions that were compared to compute indexes of holistic processing. For example, in the CFDT, we compared the upright configural to the inverted configural conditions. It may be argued that a more valid comparison would be between the upright configural and upright parts conditions. We focused on the upright versus

inverted comparison because this comparison is critical in the sense that it produces a significant difference whereas the difference between upright featural and inverted featural conditions is smaller or absent (Boutet et al., 2003; Carbon & Leder, 2005; Freire et al., 2000; Leder & Bruce, 2000). Furthermore, it has been argued that configural manipulations are too extreme to reflect naturally occurring distances between features (Taschereau-Dumouchel et al., 2010). Featural manipulations in the CFDT may also involve a combination of featural and configural changes (Riesenhuber & Wolff, 2009). Despite these limitations, we included this measure in our study because of its emphasis on sensitivity to spatial relations, which is a key concept in the literature on holistic processing and yet has not been previously investigated in individual differences studies (Maurer et al., 2002).

Finally, the results of the present study have limited implications for processing of familiar faces (Burton et al., 2015). This limitation is particularly salient for the CFDT measure because sensitivity to spatial relations appears to be less important for recognizing well-known faces (Burton et al., 2015). It would be interesting for future research to incorporate familiar faces to examine whether relationships, or lack thereof, between different measure of holistic processing of familiar faces.

Conclusion

As a whole, our results, together with those of other studies reviewed here, suggest that the four measures of holistic processing investigated herein tap into distinct, but partially overlapping, mechanisms. Hence, different measures of holistic processing should not be used interchangeably. A challenge for future research will be to develop means to dissociate between holistic processing, featural processing, and the other perceptual and cognitive mechanisms that are triggered by holistic tasks. In following with the psychometric tradition, these efforts will require a shift away from novelty and towards replication (e.g., Makel & Plucker, 2014) such that evidence can be accrued over time across different research groups. To conclude, we hope that researchers will continue to investigate this phenomenon which has captured the attention of the field for more than four decades and which is at the heart of what makes faces ‘special’.

Open practices statement This study was not preregistered. Raw accuracy and reaction time data are available at <https://osf.io/69urz/> (<https://doi.org/10.17605/OSF.IO/69URZ>).

Author note This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada. We

would like to thank Isabel Gauthier for her comments on an earlier version of this manuscript.

References

- Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Attention, Perception, & Psychophysics*, 65(7), 1078–1093. <https://doi.org/10.3758/BF03194835>
- Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition*, 34(4), 854–864. <https://doi.org/10.3758/BF03193432>
- Boutet, I., Gentes-Hawn, A., & Chaudhuri, A. (2002). The influence of attention on holistic face encoding. *Cognition*, 84(3), 321–341. [https://doi.org/10.1016/S0010-0277\(02\)00072-0](https://doi.org/10.1016/S0010-0277(02)00072-0)
- Boutet, I., Lemieux, C. L., Goulet, M. A., & Collin, C. A. (2017). Faces elicit different scanning patterns depending on task demands. *Attention, Perception, & Psychophysics*, 79(4), 1050–1063. <https://doi.org/10.3758/s13414-017-1284-y>
- Boutet, I., & Meinhardt-Injac, B. (2018). Age differences in face processing: The role of perceptual degradation and holistic processing. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 74(6), 933–942. <https://doi.org/10.1093/geronb/gbx172>
- Boutet, I., Shah, D. K., Collin, C. A., Berti, S., Persike, M., & Meinhardt-Injac, B. (2020). Age-related changes in amplitude, latency and specialization of ERP responses to faces and watches. *Aging, Neuropsychology, and Cognition*, 1–28. Advance online publication. <https://doi.org/10.1080/13825585.2019.1708253>
- Brecht, K. F., Wagener, L., Ostojić, L., Clayton, N. S., & Nieder, A. (2017). Comparing the face inversion effect in crows and humans. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 203(12), 1017–1027. <https://doi.org/10.1007/s00359-017-1211-7>
- Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 10(4), 482–496. <https://doi.org/10.1177/1745691615583129>
- Caharel, S., Fiori, N., Bernard, C., Lalonde, R., & Rebaï, M. (2006). The effects of inversion and eye displacements of familiar and unknown faces on early and late-stage ERPs. *International Journal of Psychophysiology*, 62(1), 141–151. <https://doi.org/10.1016/j.ijpsycho.2006.03.002>
- Callender, J. C., & Osburn, H. G. (1979). An empirical comparison of coefficient alpha, Guttman's lambda-2, and MSPLIT maximized split-half reliability estimates. *Journal of Educational Measurement*, 89–99.
- Carbon, C. C., & Leder, H. (2005). When feature information comes first! Early processing of inverted faces. *Perception*, 34(9), 1117–1134. <https://doi.org/10.1068/p5192>
- Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in Psychology*, 8, 1–16. <https://doi.org/10.3389/fpsyg.2017.01482>
- Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: Current perspectives. *Eye and Brain*, 8, 165–175. <https://doi.org/10.2147/EB.S92838>
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9. <http://www.pareonline.net/pdf/v10n7.pdf>
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671–684. <https://doi.org/10.1037/h0043943>
- DeGutis, J., Cohan, S., Mercado, R. J., Wilmer, J., & Nakayama, K. (2013). Holistic processing of the mouth but not the eyes in developmental prosopagnosia. *Cognitive Neuropsychology*, 29(5/6), 419–446. <https://doi.org/10.1080/02643294.2012.754745>
- DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, 126(1), 87–100. <https://doi.org/10.1016/j.cognition.2012.09.004>
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107–117. <https://doi.org/10.1037/0096-3445.115.2.107>
- Edelman, G. M., & Gally, J. A. (2013). Reentry: A key mechanism for integration of brain function. *Frontiers in Integrative Neuroscience*, 7, 1–6. <https://doi.org/10.3389/fnint.2013.00063>
- Farah, M. J., Levinson, K. L., & Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia*, 33(6), 661–674.
- Farah, M. J., Wilson, K. D., Tanaka, J. N., & Drain, M. (1998). What is “special” about face perception? *Psychological Review*, 105(3), 482–498. <https://doi.org/10.1037/0033-295X.105.3.482>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26(2), 105–109. <https://doi.org/10.3969/j.issn.1002-0829.2014.02.009>
- Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, 29(2), 159–170. <https://doi.org/10.1068/p3012>
- Gauthier, I., Chua, K.-W., & Richler, J. J. (2018). How holistic processing of faces relates to cognitive control and intelligence. *Attention, Perception & Psychophysics*, 80(6), 1449–1460. <https://doi.org/10.3758/s13414-018-1518-7>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1), 59–67. [https://doi.org/10.1016/S0006-3223\(01\)01330-0](https://doi.org/10.1016/S0006-3223(01)01330-0)
- Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, 25(4), 794–810. <https://doi.org/10.1037/a0019774>
- Hogg, R. V. (1979). An introduction to robust estimation. in R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 1–17). Academic Press.
- Huber, P. J. (1981). *Robust statistics*. John Wiley and Sons.
- Humphreys, L. G. (1990). Erroneous interpretation of difference scores: Application to a recent example. *Intelligence*, 14(2), 231–233. [https://doi.org/10.1016/0160-2896\(90\)2990005-E](https://doi.org/10.1016/0160-2896(90)2990005-E)
- Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision*, 9(6), 8.1–16. <https://doi.org/10.1167/9.6.8>
- Jaworska, K., Yi, F., Ince, R. A. A., van Rijsbergen, N. J., Schyns, P. G., & Rousselet, G. A. (2019). Healthy aging delays the neural processing of face features relevant for behavior by 40 ms. *Human Brain Mapping*, 41(5), 1212–1225. <https://doi.org/10.1002/hbm.24869>

- Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, 21(1), 38–43. <https://doi.org/10.1177/0956797609356508>
- Lavallée, M. M., Gandini, D., Rouleau, I., Vallet, G. T., Joannette, M., Kergoat, M.-J., Busigny, T., Rossion, B., & Joubert, S. (2016). A qualitative impairment in face perception in Alzheimer's disease: Evidence from a reduced face inversion effect. *Journal of Alzheimer's Disease*, 51(4), 1225–1236. <https://doi.org/10.3233/JAD-151027>
- Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology: Section A*, 53(2), 513–536. <https://doi.org/10.1080/027249800390583>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189X14545513>
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260. <https://doi.org/10.1016/S1364-6613%2802%2901903-4>
- McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, 69, 10–22. <https://doi.org/10.1016/j.visres.2012.07.014>
- McKone, E., Davies, A. A., Darke, H., Crookes, K., Wickramariyaratne, T., Zappia, S., Fiorentini, C., Favelle, S., Broughton, M., & Fernando, D. (2013). Importance of the inverted control in measuring holistic face processing with the composite effect and part-whole effect. *Frontiers in Psychology*, 4, 1–33. <https://doi.org/10.3389/fpsyg.2013.00033>
- McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1), 8–15. <https://doi.org/10.1016/j.tics.2006.11.002>
- McKone E. M., & Robbins, R. R. (2011) Are faces special? In A. C. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The handbook of face perception* (pp. 149–176). Oxford University Press.
- Meinhardt, G., Meinhardt-Injac, B., & Persike, M. (2019). Orientation-invariance of individual differences in three face processing tasks. *Royal Society Open Science*, 6(1), 1–22. <https://doi.org/10.1098/rsos.181350>
- Meinhardt-Injac, B., Boutet, I., Persike, M., Meinhardt, G., & Imhof, M. (2016). From development to aging: Holistic face perception in children, younger and older adults. *Cognition*, 158, 134–146. <https://doi.org/10.1016/j.cognition.2016.10.020>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLOS ONE*, 7(6), 1–6. <https://doi.org/10.1371/journal.pone.0038493>
- Negrini, M., Brkić, D., Pizzamiglio, S., Premoli, I., & Rivolta, D. (2017). Neurophysiological correlates of featural and spacing processing for face and nonface stimuli. *Frontiers in Psychology*, 8, 1–9. <https://doi.org/10.3389/fpsyg.2017.00333>
- Nelson, E.A., Boutet, I., & Collin, C. A. (2016). Commonly used face recognition tasks do not measure the same construct: Evidence from four holistic/configural tasks and the Cambridge Face Memory Test. *Canadian Journal of Experimental Psychology-Revue Canadienne de Psychologie Experimentale*, 70, 399–399.
- Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, 12(2), 131–146.
- Peters, J. C., Goebel, R., & Goffaux, V. (2018). From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biological Psychology*, 138, 1–10. <https://doi.org/10.1016/j.biopsycho.2018.07.009>
- Racca, A., Amadei, E., Ligout, S., Guo, K., Meints, K., & Mills, D. (2010). Discrimination of human and dog faces and inversion responses in domestic dogs (*Canis familiaris*). *Animal Cognition*, 13(3), 525–533. <https://doi.org/10.1007/s10071-009-0303-3>
- Rakover, S. S. (2013). Explaining the face-inversion effect: The face-scheme incompatibility (FSI) model. *Psychonomic Bulletin & Review*, 20(4), 665–692. <https://doi.org/10.3758/s13423-013-0388-1>
- Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology. Human Perception and Performance*, 43(12), 1961–1973. <https://doi.org/10.1037/xhp0000400>
- Richler, J. J., Bukach, C. M., & Gauthier, I. (2009a). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception, & Psychophysics*, 71(3), 530–540. <https://doi.org/10.3758/APP.71.3.530>
- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, 22(4), 464–471. <https://doi.org/10.1177/0956797611401753>
- Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic face processing. *Journal of Vision*, 14(11):10, 1–14. <https://doi.org/10.1167/14.11.10>
- Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, 15(9), 1–12. <https://doi.org/10.1167/15.9.15>
- Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, 140(5), 1281–1302. <https://doi.org/10.1037/a0037004>
- Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009b). Holistic processing of faces happens at a glance. *Vision Research*, 49(23), 2856–2861. <https://doi.org/10.1016/j.visres.2009.08.025>
- Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, 51(3), 333–342. <https://doi.org/10.1016/j.visres.2010.11.014>
- Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology*, 3, 1–6. <https://doi.org/10.3389/fpsyg.2012.00553>
- Riesenhuber, M., & Wolff, B. S. (2009). Task effects, performance levels, features, configurations, and holistic face processing: A reply to Rossion. *Acta Psychologica*, 132(3), 286–292. <https://doi.org/10.1016/j.actpsy.2009.07.004>
- Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces?. *Cognition*, 88(1), 79–107.
- Robotham, R. J., & Starrfelt, R. (2018). Tests of whole upright face processing in prosopagnosia: A literature review. *Neuropsychologia*, 121, 106–121. <https://doi.org/10.1016/j.neuropsychologia.2018.10.018>
- Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite-task measurements of holistic face processing. *Behavior Research Methods*, 47(3), 736–743. <https://doi.org/10.3758/s13428-014-0497-4>
- Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, 21, 139–253. <https://doi.org/10.1080/13506285.2013.772929>
- Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, 8(4), 1–13. <https://doi.org/10.1167/8.4.3>

- Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience*, 19(9), 1435–1452. <https://doi.org/10.1162/jocn.2007.19.9.1435>
- Schwanger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. In: H. H. Bülthoff, C. Wallraven, S.-W. Lee, & T. A. Poggio (Eds.), *Biologically motivated computer vision* (Vol. 2525, pp. 643–650). https://doi.org/10.1007/3-540-36181-2_64
- Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75(2), 221–242. <https://doi.org/10.1111/j.2044-8295.1984.tb01895.x>
- Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition*, 125(1), 64–79. <https://doi.org/10.1016/j.cognition.2012.06.005>
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1–25. <https://doi.org/10.1016/j.neulet.2005.09.001>
- Sugase-Miyamoto, Y., Matsumoto, N., & Kawano, K. (2011). Role of temporal processing stages by inferior temporal neurons in facial recognition. *Frontiers in Psychology*, 2, 1–8. <https://doi.org/10.3389/fpsyg.2011.00141>
- Sunday, M. A., Richler, J. J., & Gauthier, I. (2017). Limited evidence of individual differences in holistic processing in different versions of the part-whole paradigm. *Attention, Perception, & Psychophysics*, 79(5), 1453–1465. <https://doi.org/10.3758/s13414-017-1311-z>
- Susilo, T., Rezlescu, C., & Duchaine, B. (2013). The composite effect for inverted faces is reliable at large sample sizes and requires the basic face configuration. *Journal of Vision*, 13(13), 1–9. <https://doi.org/10.1167/13.13.14>
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Allyn & Bacon.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46(2), 225–245. <https://doi.org/10.1080/14640749308401045>
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25(5), 583–592. <https://doi.org/10.3758/BF03211301>
- Tanaka, J. W., & Simonyi, D. (2016). The “parts and wholes” of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, 69(10), 1876–1889. <https://doi.org/10.1080/17470218.2016.1146780>
- Taschereau-Dumouchel, V., Rossion, B., Schyns, P. G., & Gosselin, F. (2010). Interattribute distances do not represent the identity of real world faces. *Frontiers in Psychology*, 1, 1–10.
- Tomonaga, M. (2018). *Inverted face with upright body: evidence for face inversion effect in Japanese macaques (Macaca fuscata) under the preferential looking procedure*. BioRxiv, 1–13. <https://doi.org/10.1101/266676>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 43(2), 161–204. <https://doi.org/10.1080/14640749108400966>
- Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 69(10), 1996–2019. <https://doi.org/10.1080/17470218.2014.990392>
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, 49(2), 653–673. <https://doi.org/10.3758/s13428-016-0721-5>
- Wang, M.-Y., & Takeuchi, H. (2017). Individual recognition and the “face inversion effect” in medaka fish (*Oryzias latipes*). *eLife*, 6, 1–18. <https://doi.org/10.7554/eLife.24728>
- Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, 23(2), 169–177. <https://doi.org/10.1371/journal.pone.0058253>
- Wang, Z., Quinn, P. C., Jin, H., Sun, Y. H. P., Tanaka, J. W., Pascalis, O., & Lee, K. (2019). A regional composite-face effect for species-specific recognition: Upper and lower halves play different roles in holistic processing of monkey faces. *Vision Research*, 157, 89–96. <https://doi.org/10.1016/j.visres.2018.03.004>
- Watson, T. L. (2013). Implications of holistic face processing in autism and schizophrenia. *Frontiers in Psychology*, 4, 1–11. <https://doi.org/10.3389/fpsyg.2013.00414>
- Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of Personality and Social Psychology*, 99(3), 530–548.
- Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, 21(6), 561–579. <https://doi.org/10.1163/156856808786451408>
- Winne, P. H., & Belfry, M. J. (1982). Interpretive problems when correcting for attenuation. *Journal of Educational Measurement*, 19(2), 125–134. <https://doi.org/10.1111/j.1745-3984.1982.tb00121.x>
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141–145.
- Yong, A. G., & Pearce, S. (2013). A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), 79–94. <https://doi.org/10.20982/tqmp.09.2.p079>
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6), 747–759. <https://doi.org/10.1068/p160747n>
- Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, 15(24), 2256–2262.
- Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, 8, 1–9. <https://doi.org/10.3389/fnhum.2014.00562>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.