About-face on face recognition ability and holistic processing

Jennifer J. Richler — Department of Psychology, Vanderbilt University, Nashville, TN, USA ✉

R. Jackie Floyd — Department of Psychology, Vanderbilt University, Nashville, TN, USA ✉

Isabel Gauthier — Department of Psychology, Vanderbilt University, Nashville, TN, USA ✉

Previous work found a small but significant relationship between holistic processing measured with the composite task and face recognition ability measured by the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). Surprisingly, recent work using a different measure of holistic processing (Vanderbilt Holistic Face Processing Test [VHPT-F]; Richler, Floyd, & Gauthier, 2014) and a larger sample found no evidence for such a relationship. In Experiment 1 we replicate this unexpected result, finding no relationship between holistic processing (VHPT-F) and face recognition ability (CFMT). A key difference between the VHPT-F and other holistic processing measures is that unique face parts are used on each trial in the VHPT-F, unlike in other tasks where a small set of face parts repeat across the experiment. In Experiment 2, we test the hypothesis that correlations between the CFMT and holistic processing tasks are driven by stimulus repetition that allows for learning during the composite task. Consistent with our predictions, CFMT performance was correlated with holistic processing in the composite task when a small set of face parts repeated over trials, but not when face parts did not repeat. A meta-analysis confirms that relationships between the CFMT and holistic processing depend on stimulus repetition. These results raise important questions about what is being measured by the CFMT, and challenge current assumptions about why faces are processed holistically.

Introduction

Between 2010 and 2012, *Psychological Science* published three papers from different laboratories on the relationship between face recognition ability and holistic processing (Konar, Bennett, & Sekuler, 2010; Richler, Cheung, & Gauthier, 2011a; Wang, Li, Fang, Tian, & Liu, 2012). Holistic processing refers to the fact that, unlike objects, faces are processed in terms of wholes rather than as a collection of parts or features. Despite the central role of holistic processing in the literature on face perception and recognition (see Richler & Gauthier, 2014, and Rossion, 2013, for reviews), these papers were the first to empirically test the assumption on which this literature is built—that faces are processed holistically because holistic processing is useful for face recognition. Presumably holistic processing facilitates creation of robust face representations that can be used for recognition across changes in viewpoint and expression (Calder, Young, Keane, & Dean, 2000; McKone, 2008). It follows that those who process faces the most holistically should be the best at face recognition.

Consistent with this assumption, previous work found a small but significant correlation between holistic processing and face recognition ($r = 0.2$–0.4; DeGutis, Wilmer, Mercado, & Cohan, 2013; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Richler et al., 2011a).[1] In these studies, face recognition ability was measured by the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), and holistic processing was measured with the composite task (Farah, Wilson, Drain, & Tanaka, 1998; Young, Hellawell, & Hay, 1987). The CFMT has become a gold standard for quantifying individual differences in face recognition: It has high internal consistency (>0.8; Duchaine & Nakayama, 2006), captures a highly variable ability in the normal population (Russell, Duchaine, & Nakayama, 2009), and CFMT scores separate individuals with developmental prosopagnosia from controls (e.g., Bowles et al., 2009; Duchaine & Nakayama, 2006; Garrido, Nakayama, & Duchaine, 2008), implying construct validity, as this deficit is defined by an inability to recognize faces. In contrast,

the composite task was initially designed for use in group studies, and this measure of holistic processing is very robust in that context (see Richler & Gauthier, 2014, for a meta-analysis). It captures important differences between faces and nonface objects in novices (Farah et al., 1998; Meinhardt, Meinhardt-Injac, & Persike, 2014; Richler, Mack, Palmeri, & Gauthier, 2011), and when measured with objects, it has shown increases associated with the acquisition of perceptual expertise (e.g., Gauthier, Williams, Tarr, & Tanaka, 1998; Wong, Palmeri, & Gauthier, 2009). However, for the purpose of individual differences analyses, it has very low internal consistency (~0.2; DeGutis et al., 2013; Ross, Richler, & Gauthier, 2015).

Correlations are limited by the internal consistency (a type of reliability) of the measures. Practically, this means that while a relationship was detected between holistic processing and face recognition, these results are inconclusive about the magnitude of this relationship: The low correlation between face recognition and holistic processing could indicate that holistic processing only plays a small role in face recognition, or the observed correlations could be low because of measurement error in the base scores. As the field of individual differences in face recognition continues to grow (see Yovel, Wilmer, & Duchaine, 2014, for a review), the ability to measure the magnitude of relationships with precision will become increasingly important.

To address this issue, we designed a new measure of holistic processing, the Vanderbilt Holistic Face Processing Test (VHPT-F; Richler, Floyd, & Gauthier, 2014). The VHPT-F has higher reliability (~0.5) than the composite task, measures a very large effect size for holistic processing ($\eta^2_p = 0.75$), and is normally distributed in a normal adult population (Richler et al., 2014). Holistic processing in the VHPT-F and composite task are correlated (Wang, Ross, Gauthier, & Richler, 2015), suggesting that the VHPT-F measures the same ability as the composite task, in a more reliable manner. Surprisingly, however, while prior work using the composite task yielded a small but significant relationship with the CFMT, the correlation between holistic processing measured with the VHPT-F and the CFMT was not significant and in the opposite direction than theoretically predicted ($r = -0.17$, $n = 97$; Richler et al., 2014).

This result is theoretically unexpected: The assumption in the literature is that holistic processing facilitates extraction of the kind of information required to identify faces because they have highly similar features in the same general configuration. This result is also practically unexpected: In that first study, the VHPT-F achieved higher reliability than the composite task typically shows, and so it should have been more sensitive. We approached this puzzle by asking two questions: How do the VHPT-F and composite task differ, and what is the ability being measured by the CFMT?

In the composite task, participants are asked to judge whether one half of two sequentially presented composites (faces made from the top of one face and the bottom of another face) is the same or different while ignoring the other face half. Despite instructions to selectively attend, participants are unable to do so; the to-be-ignored face half influences performance because faces are obligatorily processed as wholes. The VHPT-F is modeled after the composite task and is based on the same operational definition, but uses a three-alternative forced choice design and includes more variety in the size of target and to-be-ignored face segments.

More importantly, in the VHPT-F a unique set of faces is used to create composites in each trial, while in the composite task, as in many cognitive tasks, face parts are repeated many times across trials (e.g., Richler, Mack, et al., 2011; Ross et al., 2015). Repetition can introduce spurious contributions in the measure, such as the ability to learn from repeated presentations and sensitivity to proactive interference (Underwood, 1957). These may contaminate the measurement of holistic processing. For example, in the composite task participants have to determine if the target part of the test face was seen in the preceding study face (requiring a "same" response), or on other trials earlier in the experiment (requiring a "different" response). Another possible effect of using a small set of stimuli is that it could actually lead participants to adopt strategies that are not viable with a larger set, thus changing what the test measures (for example, focusing attention on features that are particularly diagnostic to distinguish among a set of repeating face parts).

Stimulus repetition influences experimental measures in several domains (e.g., Endress & Potter, 2014; Malley & Strayer, 1995), and it may complicate other measures besides the composite task. For instance, the CFMT is a face learning task by design, in part to eliminate simultaneous feature matching (Duchaine & Nakayama, 2006). Participants learn six target faces, and must then indicate on each trial which of three faces matches the identity of any of the six target faces across changes in lighting and/or viewpoint. Distractor faces also repeat on the CFMT, so participants need to distinguish between recognizing a learned target and familiarity from having seen distractors on previous trials. More importantly, the ability to *recognize* faces in the experiment depends on how well they are *learned*. The CFMT is a test of face memory designed to improve the testing of face recognition, but the prior measures it replaced did not have such a long-term learning component. The intention to prevent feature

matching may have led to an assumption that it should tap into holistic processing, and the hardest trials on the CFMT show faces in visual noise in attempt to "force increased reliance on the special mechanisms that face recognition normally depends on" (Duchaine & Nakayama, 2006). But, the actual strategy used on the CFMT has not been tested.

One possibility is that the face learning ability measured by the CFMT is indeed the main ability relevant to face processing; the learning component of the task may simply facilitate measurement without confounding it. Another possibility is that the face learning ability measured by the CFMT is most relevant in situations where only a handful of faces need to be discriminated repeatedly, reducing the relevant dimensions and promoting a part-based strategy. The current body of work that speaks to the usefulness of the ability measured by the CFMT does not distinguish between these possibilities. The ability measured by the CFMT is highly heritable (Wilmer et al., 2010), highly variable in the normal population (Russell et al., 2009), and has a more extended developmental time course than many other abilities (Germine, Cashdollar, Duzel, & Duchaine, 2011), but these results do not speak to what the CFMT measures. CFMT scores separate prosopagnosic patients from controls (e.g., Bowles et al., 2009; Duchaine & Nakayama, 2006; Garrido et al., 2008), which highlights the clinical relevance of the ability it measures, but our understanding of the root problem in congenital prosopagnosia is still poor (Behrmann & Avidan, 2005); several aspects of face perception may be deficient in these patients. CFMT scores are only minimally correlated with nonface object recognition performance (Dennett, McKone, Tavashmi, et al., 2012; Wilhelm et al., 2010), which may suggest that the CFMT measures more than part processing, which is known to be important for object recognition. However, performance on eight reliable visual learning tasks (similar in format to the CFMT; McGugin et al., 2012) with nonface objects also revealed minimal intercorrelations (Gauthier et al., 2014). This suggests that such learning tests are highly domain-specific regardless of category, and that the low correlation between face and object recognition is not so telling about strategy. The CFMT is correlated ($r = 0.6$) with a test of face perception (CMFT; Duchaine, Yovel, & Nakayama, 2007), where subjects order simultaneously presented faces based on similarity to a target face (Bowles et al., 2009). Although this task discourages local feature use, it requires selective attention to a single dimension (related to the target face) created by morphing two faces (Goldstone & Steyvers, 2001), task demands known to allow subjects to ignore other dimensions, perhaps discouraging holistic processing. Under what circumstances holistic processing is useful for a face

recognition test, and how correlations between tasks are constrained by available strategies, need to be evaluated directly.

Given that stimulus repetition is a salient difference between the composite task and VHPT-F measures of holistic processing, and given the learning component in the CFMT, we hypothesized that learning in the CFMT may account for all of the small amount of shared variance with the relatively unreliable composite task, where face parts repeat (e.g., DeGutis et al., 2013; McGugin et al., 2012; Richler et al., 2011a), but not with the more reliable VHPT-F, where face parts do not repeat (Richler et al., 2014).

In Experiment 1, we replicated our prior finding of a lack of correlation between face recognition ability measured with the CFMT and holistic processing measured in the VHPT-F. In Experiment 2, we test the possibility that when observed, correlations between holistic processing measures and the CFMT are driven by this unforeseen contribution of learning processes related to stimulus repetition in both tasks.

Experiment 1

Methods

Participants

In Experiment 1, 119 Vanderbilt University undergraduates (38 males; mean age = 18.8 years) participated in exchange for course credit. With this sample, we had 0.8 power[2] to detect a correlation of $r = 0.25$, which is smaller than the correlation we observed with the composite task in prior work ($r = 0.40$; Richler et al., 2011a).

Procedure

All participants completed the CFMT long form, the VHPT-F, and the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). Effects related to the PANAS did not replicate and are not discussed further.

CFMT long form

In the CFMT, participants complete an 18-trial introductory learning phase, in which a target is presented in three views, followed by three forced-choice test displays containing the target face and two distractor faces. Participants are told to select the face that matches the previously studied target. This is repeated for all six target faces. Then, participants study frontal views of all six target faces together for a total of 20 s, followed by 30 forced-choice test displays.

Participants are told to select the face that matches one of the original six target faces. The matching faces vary from the studied versions in terms of lighting condition, pose, or both. Next, participants are given another opportunity to study the six target faces, followed by 24 test displays presented in Gaussian noise. For a complete description of the CFMT, see Duchaine and Nakayama (2006).

The CFMT long form includes 30 additional "difficult" trials where faces are shown as silhouettes, in extreme noise, or with varying expressions.

Vanderbilt Holistic Face Processing Test

We used a modified version of the VHPT-F described in Experiment 1 of Richler et al. (2014). Only the aligned and phase-scrambled control conditions were included. Holistic processing is indexed by failures of selective attention that manifest as a congruency effect when target and task-irrelevant face segments are aligned. The phase-scrambled condition was included as a baseline to test whether inclusion of a baseline condition is useful for quantifying individual differences in holistic processing. On each trial, a study composite was presented for 2 s with the target part outlined in red, followed by a random pattern mask (500 ms). A test display containing three composite faces was then presented, and participants had to indicate which of the three composite faces contained the target part with the same identity (outlined in red) as the study composite, while ignoring the task-irrelevant distractor parts. The correct response always showed a different image of the same individual. Response keys (J, K, L) were presented underneath the test images. The test display was presented until participants made a response. On congruent trials, the target part was paired with the same distractor part as in the study composite. On incongruent trials, the target part was paired with a different distractor part at test. On phase-scrambled trials, distractor parts at study and test were phase-scrambled (see Figure 1 for example aligned and phase-scrambled trials).

There were 140 trials (36 famous face composites, 104 unfamiliar face composites; 35 trials for each combination of aligned/phase-scrambled × congruent/incongruent). Trials were blocked by target part (top 2/3, bottom 2/3, top 1/2, bottom 1/2, top 1/3, bottom 1/3, eyes, nose, mouth). The first four trials in each block showed each of the four possible trial types (aligned congruent, aligned incongruent, phase-scrambled congruent, phase-scrambled incongruent) with composites made from famous faces as practice trials. Famous face trials were not included in the analysis because individual differences in experience with the specific face identities and the ability to use their names might confound individual differences in holistic processing in
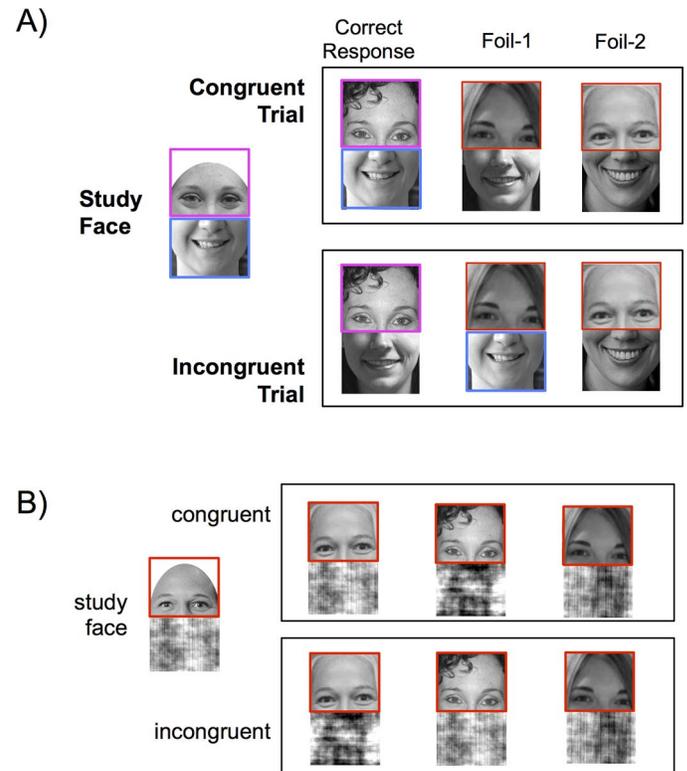

Figure 1. Example VHPT-F trials. (A) Example of aligned congruent and incongruent trials. The target and distractor segment from the study composite are shown outlined in pink and blue for illustrative purposes. (B) Phase-scrambled congruent and incongruent trials. On congruent trials, the target segment is paired with the same distractor segment as during study. On incongruent trials, the target segment is paired with a different distractor segment. Adapted from Richler et al. (2014); permissions available from the first author.

these trials. See Richler et al. (2014) for a full description of the VHPT-F.

Results

Because this experiment included the CFMT, and there is evidence that face recognition is impaired for faces from other races (e.g., Meissner & Brigham, 2001), data from non-Caucasian participants were excluded from the analyses (*n* = 3). Data from an additional seven participants were discarded due to computer errors. Therefore, data from 109 participants were analyzed.

A 2 (congruency: congruent/incongruent) × 2 (condition: aligned/phase-scrambled) repeated-measures ANOVA indicated significant holistic processing in the VHPT-F (see Figure 2): Accuracy was higher on congruent versus incongruent trials, $F(1, 108) = 169.37$, $MSE = 70.92$, $p < 0.001$, $\eta^2_p = 0.61$, and there was a significant interaction between congruency and condi-
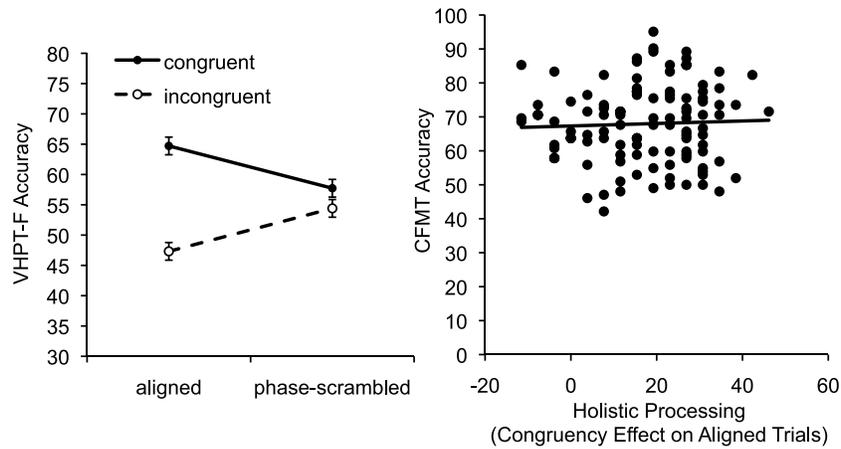
Figure 2. Left: Mean accuracy on the VHPT-F as a function of trial condition and congruency. Error bars show 95% confidence intervals of within-subjects effects. Right: Correlation between holistic processing measured in the VHPT-F (congruency effect on aligned trials; *x*-axis) and performance on the CFMT (*y*-axis).

tion, $F(1, 108) = 69.99$, $MSE = 74.12$, $p < 0.001$, $\eta^2_p = 0.39$. Mean accuracy on the CFMT was 69.32% ($SD = 11.8$).

The correlation (Pearson's *r*) between holistic processing and face recognition ability measured by the CFMT was very small and not significant, regardless of whether holistic processing was indexed by the congruency effect on aligned trials only (performance on congruent trials minus performance on incongruent trials, $r(109) = 0.042$, $p = 0.67$, CI $= -0.147, 0.228$), or the congruency × condition interaction (congruency effect on aligned trials with the congruency effect on phase-scrambled trials regressed out; $r[109] = 0.043$, $p = 0.66$, CI $= -0.146, 0.229$; see Figure 2; see Table 1 for task reliabilities and disattenuated correlations). The fact that these correlations are nearly identical is consistent with previous suggestions that, because the congruency effect in baseline conditions shares little variance with the congruency effect on aligned trials, $r(109) = -0.02$ in this dataset, baseline trials are not useful for quantifying individual differences in holistic processing (Richler et al., 2014; Richler & Gauthier, 2014).

Discussion

Experiment 1 replicated the previous finding that holistic processing measured in the VHPT-F is not correlated with face recognition ability measured by the CFMT (Richler et al., 2014). This is surprising given

that small but significant correlations have been found between the CFMT and the composite task (DeGutis et al., 2013; McGugin et al., 2012; Richler et al., 2011a), which generally provides a less reliable measure of holistic processing (DeGutis et al., 2013; Ross et al., 2014),[3] and given that theories of face perception are based on the assumption that holistic processing facilitates face recognition (e.g., Calder et al., 2000; McKone, 2008).

In Experiment 2 we test the hypothesis that correlations between the CFMT and the less reliable composite task are driven by processes recruited when faces or face parts repeat (such as learning or proactive interference), not holistic processing. To this end, participants completed the CFMT, a longer and therefore more reliable VHPT-F, and two versions of the composite task. In the composite task with few faces (Comp-Few), a small set of face top and bottom halves were used to create composites, so face parts repeated many times across trials. In the composite task with many faces (Comp-Many), a large set of face tops and bottoms were used to make composites, so part repetition was rare. Comp-Few and Comp-Many should both measure holistic processing, and so should correlate with each other and the VHPT-F. We expect to replicate the absence of a relationship between holistic processing in the VHPT-F and the CFMT. Critically, if stimulus repetition drives correlations with the CFMT, we would expect the CFMT to correlate with holistic processing measured in Comp-Few, but not Comp-Many.

Test	Reliability	Corrected correlation with CFMT
CFMT	0.92[a]	
VHPT-F (with phase-scrambled control)	0.18[b]	0.11
VHPT-F (aligned trials only)	0.19[b]	0.10

Table 1. Reliability for all tasks and disattenuated correlations. *Notes*: [a] Cronbach's alpha; [b] Guttman $\lambda 2$.

Experiment 2

Methods

Participants

Sample size was determined by a power analysis on the correlation between VHPT-F and the composite task, which is likely the most difficult correlation for us to detect because of the low reliability of the composite task. We made the following assumptions: We assumed that if the composite task and the VHPT-F tap into the same construct of holistic processing, the shared variance should be at least 25% ($r = 0.5$). Assuming reliabilities of 0.52 for the version of the longer VHPT-F used in this experiment (Richler et al., 2014) and of 0.2 for the version of the composite task for which we expected the lowest reliability (that with nonrepeated faces; Ross et al., 2015), the attenuated correlation would be 0.16 [$0.5 \times \mathrm{SQRT}(0.52 \times 0.2)$]. The sample required for 0.80 power to detect this correlation with a one-tail test (as a negative correlation is not theoretically justifiable nor is it expected given prior results) is 234 subjects.

In Experiment 2, 236 members of the Vanderbilt University community (100 male; mean age = 22.4 years) participated in exchange for monetary compensation or course credit.

Procedure

Participants completed two sessions. In Session 1, participants completed the composite task with few (Comp-Few) or many (Comp-Many) faces (order counterbalanced), followed by the CFMT long form. In Session 2, participants completed Comp-Many or Comp-Few (depending on which version was completed in Session 1), followed by the VHPT-F.

CFMT long form

The CFMT long form was the same as Experiment 1.

Vanderbilt Holistic Face Processing Test

Here we used the version of the VHPT-F described in Experiment 2 of Richler et al. (2014) that includes more aligned trials, and no misaligned or phase-scrambled baseline trials because we have repeatedly found that the congruency effect in baseline conditions does not correlate with the aligned congruency effect (Richler et al., 2014; Richler & Gauthier, 2014), and so there is no shared variance to regress out.

There were 134 trials (18 famous, 116 unfamiliar; 67 congruent/incongruent). Trials were blocked by target part (top 2/3, bottom 2/3, top 1/2, bottom 1/2, top 1/3, bottom 1/3, eyes, nose, mouth). The first two trials (one congruent, one incongruent) in each block were practice trials showing composites made from famous faces (only unfamiliar face trials were analyzed).

Composite task

Stimuli for the composite task were 100 female faces from the Max Planck Institute Database that were converted to gray-scale and cut in half to produce 100 face tops and 100 face bottoms. For each participant, five tops and five bottoms were randomly selected, with the constraint that tops and bottoms could not come from the same original faces. These five tops and bottoms were used to create composites in Comp-Few. The remaining 95 tops and bottoms were used to create composites in Comp-Many.

On each composite task trial, a fixation cross was presented (500 ms), followed by a study composite (200 ms), a random pattern mask (500 ms), and a test composite (200 ms). Participants were instructed to indicate whether the target half (top or bottom) of the test composite was the same or different as in the study composite, while ignoring the other, task-irrelevant half. On aligned trials, both halves of the study and test composite were aligned, with a white line 3 pixels thick separating the face halves, such that the study composite was 256×256 pixels. On misaligned trials, the study composite was aligned, and the test composite was misaligned (top half moved 32 pixels to the left and bottom half moved 32 pixels to the right, resulting in a test composite 320×256 pixels). Misaligned trials were included to maintain the same experimental design as the previous studies that found correlations between the CFMT and composite task.

The target half was blocked, and top and bottom blocks were alternated. There were four blocks total, and the top block was first for all participants. Composite presentation times were reduced to 150 ms (from 200 ms) in the second top and bottom blocks. We hoped that this would help reliability by adding variability in task difficulty. There were 20 trials for each combination of target part (top/bottom), alignment (aligned/misaligned), congruency (congruent/incongruent), and correct response (same/different), for a total of 320 trials. A practice block of 16 trials (eight top and eight bottom) preceded the experimental trials.

Results

Because we were explicitly recruiting Caucasian participants only, data from non-Caucasian participants ($n = 3$) were discarded. Data from four participants were discarded because they did not show up for the second session, and data from 12 partici-
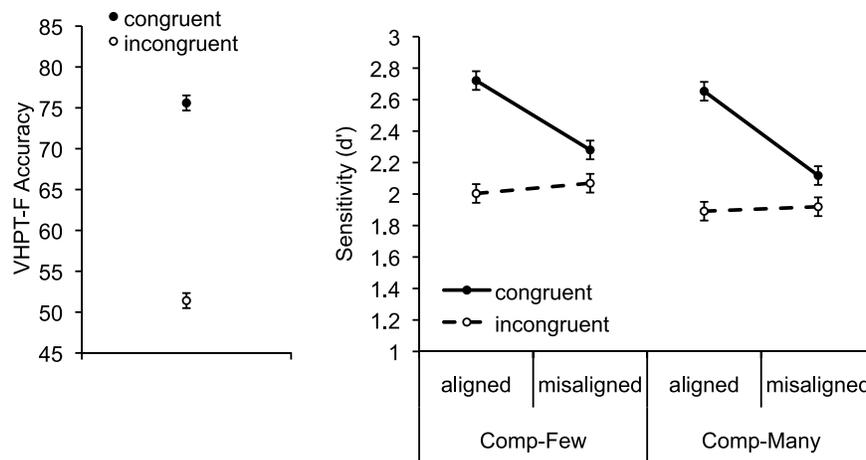
Figure 3. Left: Mean accuracy on the VHPT-F as a function of congruency. Right: Mean sensitivity (d') in Comp-Few and Comp-Many as a function of alignment and congruency. Error bars show 95% confidence intervals of within-subjects effects.

pants were discarded due to computer/experimenter errors. Next, we discarded data from participants with below chance performance (average $d' < 0$) in Comp-Many ($n = 5$) and Comp-Few ($n = 2$). Finally, data from one participant were discarded for being an extreme outlier (accuracy = 7%) on incongruent VHPT-F trials. Thus, data from 209 participants were included in the analyses (still affording statistical power of 0.76 to detect a very small correlation).

Group-level analyses

A 2 (repetition: Comp-Few/Comp-Many) × 2 (alignment: aligned/misaligned) × 2 (congruency: congruent/incongruent) repeated-measures ANOVA was conducted on d' in the composite task (see Figure 3). There were significant main effects of alignment, $F(1, 208) = 122.33$, $MSE = 0.17$, $p < 0.001$, $\eta^2_p = 0.37$, and congruency, $F(1, 208) = 443.45$, $MSE = 0.21$, $p < 0.001$, $\eta^2_p = 0.68$, and a significant congruency × alignment interaction, $F(1, 208) = 157.65$, $MSE = 0.19$, $p < 0.001$, $\eta^2_p = 0.43$. Critically, although performance was higher overall in Comp-Few versus Comp-Many, $F(1, 208) = 17.12$, $MSE = 0.37$, $p < 0.001$, $\eta^2_p = 0.08$, repetition did not interact with any other factors (ps > 0.09, $\eta^2_p <$ 0.02). In particular, repetition did not influence the congruency effect on aligned trials only (Comp-Many: $\eta^2_p = 0.58$; Comp-Few: $\eta^2_p = 0.53$) nor the congruency × alignment interaction (Comp-Many: $\eta^2_p = 0.32$; Comp-Few: $\eta^2_p = 0.26$), indicating that holistic processing was similar in magnitude across composite tasks.

Holistic processing in the VHPT-F (congruency effect on aligned trials; Richler et al., 2014; see Figure 3) was significant, $F(1, 208) = 1335.58$, $MSE = 45.78$, $p < 0.001$, $\eta^2_p = 0.87$. Mean accuracy on the CFMT was 71.82% ($SD = 10.99$).

Individual differences analyses

Correlations between all tasks are shown in Table 2. We used a version of the VHPT-F that only included aligned trials. Therefore, for consistency, holistic processing in the correlations was indexed as the congruency effect on aligned trials for Comp-Few and Comp-Many as well.[4] Because correlations may be particularly sensitive to outliers, we also examined the correlations using two outlier removal methods: skipped correlations (Pernet, Wilcox, & Rousselet, 2013) and winsorizing (Erceg-Hurn & Mirosevich, 2008). Importantly, most of the correlations, including the critical correlations between holistic processing and the CFMT, are unaffected by how outliers are handled.[5] Table 3 shows reliability for all tasks and the critical correlations between holistic processing and the CFMT corrected for reliability.

Discussion

The results of Experiment 2 are consistent with our predictions: CFMT performance was correlated with holistic processing measured in a task where face parts were repeated (Comp-Few), but not with holistic processing measured in tasks where face parts did not repeat (Comp-Many, VHPT-F). These results support our hypothesis that correlations with the CFMT are driven by processes recruited when stimuli repeat, such as learning or proactive interference, not holistic processing. Our results cannot speak to the quantitative contribution of face learning or proactive interference in this relationship because the holistic processing measures are not sufficiently reliable. Indeed, the shared variance we observed even with face repetition is small, and we simply do not have sufficient power to show that the correlation is significantly larger than the

Holistic processing measure	CFMT			Comp-Many			Comp-Few		
	r	Skipped	Winsorized	r	Skipped	Winsorized	r	Skipped	Winsorized
Comp-Many	0.06 [−0.08, 0.19]	0.06 [−0.07, 0.18]	0.06 [−0.08, 0.19]						
Comp-Few	0.15* [0.02, 0.28]	0.13* [0.01, 0.25]	0.16* [0.03, 0.29]	0.17* [0.04, 0.30]	0.23* [0.09, 0.35]	0.15* [0.02, 0.28]			
VHPT	0.01 [−0.13, 0.15]	0.02 [−0.11, 0.15]	−0.002 [−0.14, 0.13]	0.09 [−0.05, 0.22]	0.10 [−0.01, 0.23]	0.16* [0.03, 0.29]	0.23* [0.10, 0.35]	0.22* [0.10, 0.34]	0.26* [0.13, 0.38]

Table 2. Correlations (Pearson's r) between all holistic processing measures computed three ways (no outlier removal, skipped correlations, winsorized correlations). Notes: * $p < 0.05$.

			Corrected correlation with CFMT	
Task	Reliability	r	Skipped	Winsorized
CFMT	0.89[a]			
Comp-Many	0.34[b]	0.11	0.11	0.11
Comp-Few	0.38[b]	0.26	0.22	0.028
VHPT-F	0.38[c]	0.02	0.03	−0.003

Table 3. Reliability for all tasks and disattenuated correlations with CFMT. Notes: [a] Cronbach's alpha; [b] Spearman-Brown corrected split-half reliability; [c] Guttman $\lambda 2$.

correlation without repetition (but see the General discussion where we address this issue). Importantly, however, here we were interested in a qualitative prediction: The critical result is the complete absence of a correlation between CFMT and the composite task once repetition of face parts was removed.

In principle it may be possible to manipulate the composite task parameters to increase the magnitude of the relationship with the CFMT when face parts repeat, but it is unclear why this would be useful given that our results suggest that holistic processing—what the composite task is meant to measure—is not driving the relationship in the first place.

General discussion

Coming to this project, we knew of several demonstrations of a correlation between CFMT and holistic processing in the composite task where face parts repeated (DeGutis et al., 2013; McGugin et al., 2012; Richler et al., 2011a). These correlations were small, and could also have been inflated because the sample sizes were relatively small ($n = 38–66$). In Experiment 2, we again found this correlation to be significant (correlation between CFMT and Comp-Few), but the effect size was smaller than previous studies, perhaps because our larger sample ($n = 209$) provided a better estimate that was less susceptible to outliers. To get a better sense of the true magnitude of this relationship, we plotted the magnitude of the correlation between CFMT and holistic processing obtained in studies where face parts were repeated (Figure 4, gray squares). A meta-analysis weighted by sample size yielded an effect size of $r = 0.21$ (95% CI = 0.11, 0.31, black square). Thus, these results converge to show a small but significant effect.

But, is this effect due to a true contribution of holistic processing to performance on the CFMT? The goal of the present work was to reduce the contribution of stimulus repetition in holistic processing measures, and see if this relationship would survive. This was investigated using a version of the composite task with
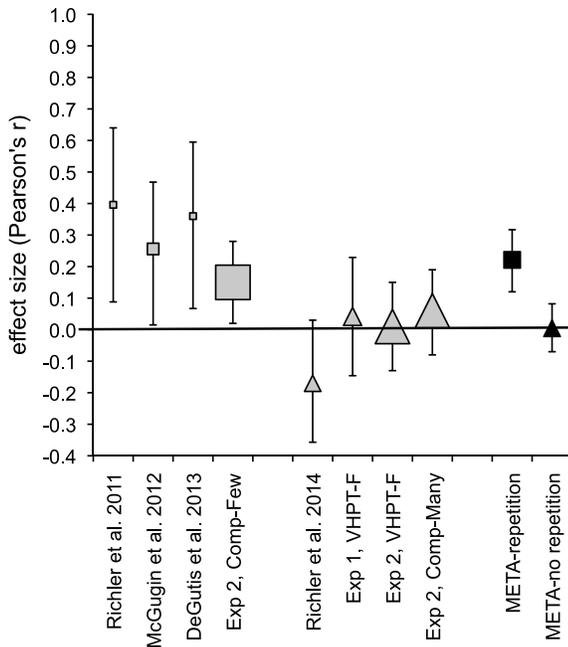
Figure 4. Forest plot of the effect size (Pearson's *r*) of the relationship between CFMT performance and holistic processing in tasks where face parts repeated (gray squares) and in tasks where face parts did not repeat (gray triangles). Marker size is proportional to weights used in the meta-analysis. Black markers show meta-analytic effect sizes. Error bars show 95% confidence intervals.

face parts that rarely repeated, and the VHPT-F where face parts do not repeat. In the group average analysis, these measures capture holistic processing as well as the composite task with repeating faces, and show comparable reliability (see Table 3).[6] However, each time we related the CFMT to holistic processing when learning specific face parts was not possible, we observed no relationship. The magnitude of these relationships is also plotted in Figure 4 (gray triangles). A meta-analysis of these studies weighted by sample size yielded an effect size of $r = 0.004$ (95% CI = −0.07, 0.08, black triangle). Thus, these results converge to show no relationship with the CFMT when face parts do not repeat in the holistic processing task.

Importantly, as can be appreciated from Figure 4, the error bars around the meta-analytic effect sizes for studies with face part repetition (black square) and studies with no face part repetition (black triangle) are nonoverlapping, and the difference between them is statistically significant ($p = 0.0009$). Therefore, it does not appear to be the case that all of these effects are from the same distribution, with greater precision in the studies where stimuli did not repeat because these studies tended to have larger samples. Rather, this supports our hypothesis that relationships between holistic processing measures and the CFMT are driven by processes recruited when face parts repeat. While

our results do not speak to whether this reflects an influence of learning diagnostic features or proactive interference, or both, the important point is that without stimulus repetition, there is no correlation between holistic processing and face recognition ability.

These results raise important questions about the CFMT and holistic processing. The CFMT has been described as capturing the real-world demands of face recognition, where the same faces must be remembered and repeatedly recognized despite superficial changes in lighting, pose, hairstyle, etc. However, the learning component of the task may promote strategies that do not tap into perceptual abilities. Indeed, studies that find correlations between the CFMT and face perception measured by face identity aftereffects also use designs where stimuli repeat (Dennett, McKone, Edwards, & Susilo, 2012; Rhodes, Jeffrey, Taylor, Hayward, & Ewing, 2014), and it remains to be seen whether this relationship will survive if face repetition is eliminated. This is not to say face perception is more important than face memory, or vice versa, but rather that researchers investigating individual differences need to more carefully consider the kind of face recognition ability reflected in their choice of task, as these may influence the processes that are recruited and the kind of strategies participants use.

It is possible that the VHPT-F will correlate with face recognition ability measured in a task like the CFMT but where faces do not repeat, and participants must rely on perceptual abilities rather than learning and memory. Of course, an alternative possibility is that holistic processing and face recognition ability are not correlated. To be clear, this would not mean that holistic processing is not involved in face and expert object perception: Faces but not objects are processed holistically (e.g., Farah et al., 1998; Meinhardt et al., 2014; Richler, Mack, et al., 2011), and increases in holistic processing for objects are observed following the acquisition of perceptual expertise (e.g., Gauthier et al., 1998; Wong et al., 2009). These effects are robust at the group level and replicate consistently. Thus, the challenge for future research is to understand why and under what circumstances faces and objects of expertise are processed holistically without assuming that holistic processing promotes good recognition ability in all situations.

Keywords: holistic processing, face recognition, individual differences

Acknowledgments

Footnotes

[1] The other two studies published in *Psychological Science* (Konar et al., 2010; Wang et al., 2012) yielded inconsistent results. However, these results are difficult to interpret because the measure of holistic processing used has been shown to track response biases unrelated to holistic processing, and the influence of response bias is fully confounded with the holistic processing measure in that task (see Cheung, Richler, Palmeri, & Gauthier, 2008; Richler et al., 2011a, 2011b, and Richler, Mack, et al., 2011, for empirical demonstrations of this problem, and Richler & Gauthier, 2014, for a review). The arguments in the present work do not affect the conclusions of our prior work about the validity of this older version of the composite task.

[2] After discarding participants as discussed in the Results, $n = 109$ and power was still 0.76.

[3] The low internal consistency of the version of the VHPT-F used in Experiment 1 is likely due to the fact that it had a small number of trials (26 trials per condition) compared to other versions of the VHPT-F (for instance, the VHPT-F 2.0 has 90 trials per condition). However, it is important to note that reliability is not a property of a task; it is a property of measurements (Thompson & Vacha-Haase, 2000), so it is also possible for identical tasks to produce different reliabilities in different samples.

[4] For both Comp-Few and Comp-Many, there was little shared variance between the congruency effect on aligned trials and that on misaligned trials ($rs < 0.07$, $ps > 0.3$). Thus, including the misaligned trials would have very little impact on individual differences measures of holistic processing.

[5] The only qualitative difference is that the correlation between VHPT-F and Comp-Many was only significant for winsorized scores. It is not clear why VHPT-F correlated with Comp-Few and not Comp-Many, but given that this difference is not significant, and the result is unexpected, it should be interpreted with caution unless it can be replicated.

[6] Although in principle the VHPT-F was designed to have better reliability than the composite task, this was not the case in this particular study. This underscores the important point that reliability is property of

measurement, not task (Thompson & Vacha-Haase, 2000).

References

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. *Trends in Cognitive Science*, *9*, 180–187.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., . . . Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 432–455.

Calder, A. J., Young, A.W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception & Performance*, *26*(2), 527–551.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception & Performance*, *34*(6), 1327–1336.

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*, 87–100.

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, *23*, 1279–1287.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*, 576–585.

Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social, Cognitive, & Affective Neuroscience*, *2*, 104–113.

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General*, *143*(2), 548–565.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, *63*(7), 591–601.

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

Garrido, L., Nakayama, K., & Duchaine, B. (2008). Face detection in normal and prosopagnosic indivdiduals. *Journal of Neuropsychology*, *2*, 119–140.

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & VanGulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision*, *14*(8):7, 1–12, doi:10.1167/14.8.7. [PubMed] [Article]

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. W. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.

Germine, L., Cashdollar, N., Duzel, E., & Duchaine, B. (2011). A new selective developmental deficit: Impaired object recognition with normal face recognition. *Cortex*, *47*, 598–607.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, *21*, 38–43.

Malley, G. B., & Strayer, D. L. (1995). Effect of stimulus repetition on positive and negative identity priming. *Perception & Psychophysics*, *57*(5), 657–667.

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.

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 310–327.

Meinhardt, G., Meinhardt-Injac, B., & Persike, M. (2014). The complete design in the composite face paradigm: Role of response bias, target certainty, and feedback. *Frontiers in Human Neuroscience*, *8*, 885.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law*, *7*, 3–35.

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, *3*, 606.

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., & Ewing, L. (2014). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *Journal of Experimental Psychology: Human Perception & Performance*, *40*, 897–903.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological Science*, *22*, 464–471.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011b). Beliefs alter holistic face processing. . .if response bias is not taken into account. *Journal of Vision*, *11*(13):17, 1–13, doi:10.1167/11.13.17. [PubMed] [Article]

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, doi:10.1167/14.11.10, doi:10.1167/11.13.17. [PubMed] [Article]

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic processing. *Psychological Bulletin*, *140*(5), 1281–1302.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*, 333–342.

Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite task measurements of holistic face processing. *Behavior Research Methods*, in press.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*, 139–253.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational & Psychological Measurement*, *60*(2), 174–195.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60.

Wang, C.-C., Ross, D. A., Gauthier, I., & Richler, J. J. (2015). Validation of the Vanderbilt Holistic Face Processing Test. *Manuscript in preparation*.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012).

Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23,* 169–177.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality & Social Psychology, 54*(6), 1063–1070.

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 & Social Psychology, 99*(3), 530–548.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., . . . Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, 107,* 5238–5241.

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert—But which type? *Psychological Science, 20,* 1108–1117.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747–759.

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience, 8,* 562.