CrossMark

Limited evidence of individual differences in holistic processing in different versions of the part-whole paradigm

Mackenzie A. Sunday[1] · Jennifer J. Richler[1] · Isabel Gauthier[1]

Abstract The part-whole paradigm was one of the first measures of holistic processing and it has been used to address several topics in face recognition, including its development, other-race effects, and more recently, whether holistic processing is correlated with face recognition ability. However the task was not designed to measure individual differences and it has produced measurements with low reliability. We created a new holistic processing test designed to measure individual differences based on the part-whole paradigm, the Vanderbilt Part Whole Test (VPWT). Measurements in the part and whole conditions were reliable, but, surprisingly, there was no evidence for reliable individual differences in the part-whole index (how well a person can take advantage of a face part presented within a whole face context compared to the part presented without a whole face) because part and whole conditions were strongly correlated. The same result was obtained in a version of the original part-whole task that was modified to increase its reliability. Controlling for object recognition ability, we found that variance in the whole condition does not predict any additional variance in face recognition over what is already predicted by performance in the part condition.

Keywords Part-whole paradigm · Holistic processing · Face recognition

Several phenomena suggest that faces are processed as singular units rather than by individual facial features. Such holistic processing is considered a hallmark of face perception (e.g., Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993; but see Donnelly, Cornes, & Menneer, 2012; Fitousi, 2015), and yet its role in face recognition is still not fully understood. In particular, one assumption is that holistic processing contributes to the efficacy of face recognition (Richler & Gauthier, 2014; Rossion, 2013), such that those who excel at face recognition should rely more heavily on this kind of processing. Interestingly, the best evidence linking recognition ability to holistic processing is obtained with non-face objects, whereby holistic processing increases with perceptual expertise (Chua, Richler, & Gauthier, 2015; Gauthier, Williams, Tarr, & Tanaka, 1998; Wong, Palmeri, & Gauthier, 2009). These studies with objects used a holistic processing measure, the composite paradigm, which defines holistic processing as a failure of selective attention to parts (i.e., participants are instructed to attend to only part of the object but are unable to do so and consequently are influenced by the other parts). When this paradigm is used with faces, holistic processing is detected with a large effect size, compared to negligible effects for objects in novices (Richler & Gauthier, 2014). However, in terms of the relation between holistic effects and face recognition ability, when confounds from stimulus repetition in the holistic processing measure are removed, holistic processing of faces does not predict face recognition ability (Richler, Floyd, & Gauthier, 2015), though there is some evidence of impaired holistic processing in individuals with congenital prosospagnosia (Avidan, Tanzer, & Behrmann, 2011; Carbon, Grüter, Weber, & Lueschow, 2007).

Nevertheless, because holistic processing is a central construct in the study of face recognition (e.g., Maurer & Young, 1983; Richler, Palmeri, & Gauthier, 2012), and because it is associated with non-face expertise (Boggan, Bartlett, & Krawczyk, 2012; Bukach, Philips, & Gauthier, 2010; Busey & Parada, 2010), it is difficult to abandon the idea that people

✉ Mackenzie A. Sunday
mackenzie.a.sunday@vanderbilt.edu

1   Department of Psychology, Vanderbilt University, 226 Wilson Hall, Nashville, TN 37204, USA

Springer

who are better at face recognition process faces more holistically. So far, only one measure of holistic processing (the composite task, Richler, Floyd, & Gauthier, 2014) has been adapted for the study of individual differences. Because research on individual differences benefits from measuring abilities as latent variables that capture the shared variance among multiple indicators of a construct (Bollen, 2002), it is important to provide other ways to reliably measure individual differences in face-specific holistic processing. Therefore, here we aimed to adapt another paradigm, the part-whole task, for this purpose. The underlying motivation was to ask whether individual differences in holistic processing as measured by a different paradigm might predict face recognition ability. As we discuss below, there is no clear empirical evidence that the part-whole task taps into the same meaning of holistic processing as the composite task (Richler et al., 2012), and so their respective relations to face recognition ability could differ.

Because research on holistic processing has historically focused on group-level effects, many of these tests excel at capturing group effects (Richler & Gauthier, 2014), but lack the reliability necessary to measure individual differences (e.g., DeGutis, Wilmer, Mercado, & Cohan, 2013; Ross, Richler, & Gauthier, 2015). Individual differences work requires tests that produce reliable scores, which at a minimum means that the tests have good internal consistency. Put simply, a measure that does not correlate with itself cannot be expected to correlate with other measures. Currently, only one holistic processing test, the Vanderbilt Holistic Processing Test – Face (VHPT-F; Richler et al., 2014), was designed specifically to measure individual differences. Although far from perfectly reliable, it has produced more reliable measurements (~.6) than the task from experimental studies on which it is based (~.2–.4, DeGutis, Wilmer, et al., 2013; Ross et al., 2015). A high score on the VHPT-F reflects an inability to selectively attend to single face parts that are clearly identified on each trial. Despite reliable variability between individuals in their ability to selectively attend to face parts on this test, this variability appears to be unrelated to face recognition, at least as measured by the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), a popular measure of face recognition ability (Richler et al., 2015).

Although the definition of holistic processing targeted by the VHPT-F is not related to CFMT performance, other definitions of holistic processing could be. For instance, an individual's ability to use information from a whole face when it is available may be more relevant to face recognition than an individual's ability to selectively attend to face parts. In the part-whole paradigm (Tanaka & Farah, 1993), face parts are presented alone or in the context of complementing face parts (i.e., in a whole face) that do not add diagnostic information. The rationale behind this paradigm is that the whole context changes the manner in which parts are encoded, such that a holistic process is only engaged when face parts are shown in a whole face context (DeGutis, Mercado, Wilmer, & Rosenblatt, 2013). Although previous studies have examined how part-whole and face recognition measures relate, the part-whole measurements had relatively low reliabilities, limiting correlations with other measures (e.g., $\lambda_2 = .31$, DeGutis, Wilmer, et al., 2013; $\lambda_2 = .33$, DeGutis, Mercado, et al., 2013). In addition, in these studies a small number of faces were repeated in both the face recognition task and the part-whole task, which can inflate correlations between tasks (Richler et al., 2015). Therefore, to investigate if face recognition ability and the part-whole effect relate, we first needed to create a new version of the part-whole paradigm that produces reliable measurements and has limited stimulus repetition.

Overview

In Study 1, we created a new part-whole test designed to reliably measure individual differences in holistic processing (the Vanderbilt Part Whole Test, VPWT). In Studies 2 and 3 we test if one aspect of our new task, placing isolated parts in the context of phase-scrambled noise, produces different results from the more traditional presentation of face parts in isolation, and compare the VPWT to the original part-whole paradigm to ensure that our results are not test specific. Finally, in Study 4 we examine how VPWT performance relates measures of face (CFMT) and object (Vanderbilt Expertise Test; VET) recognition abilities.

Study 1

We made a number of modifications to the original part-whole paradigm in attempts to maximize its validity and reliability,[1] which are discussed below.

Methods

Participants

Several pilot tests were run on Amazon Mechanical Turk to develop the VPWT 1.0 ($N = 20$–137 per pilot test, total $N = 568$) and participants were paid between US$0.45 and US$0.65 for completing one test. When used properly, online crowdsourcing tools like Amazon Mechanical Turk can provide high-quality data (Buhrmester, Kwang, & Gosling, 2011; Hauser & Schwarz, 2016; Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010), and thus are a good choice for task development and piloting given the ease and speed at which data can be collected. Furthermore, one study has

---

[1] To preempt concerns about these changes, it is useful to note that in Study 3 we replicate our main results with a version of the original task.

reported only a negligible difference between online and lab participant pools for individual differences measurements (Cho et al., 2015).

Data for the VPWT 1.0 were collected as part of a larger dataset that also included the Vanderbilt Face Matching Test, and VHPT-F. Only the data from the VPWT 1.0 are reported here. One hundred and sixty-four participants were recruited from Amazon Mechanical Turk to complete the Vanderbilt Face Matching Test. We contacted these participants one day after they completed the Vanderbilt Face Matching Test (VFMT) to offer them the opportunity to complete the VPWT and VHPT-F. Participants were compensated US$0.85 for completing the VFMT, and given an additional bonus of US$2.50 if they completed both the VPWT and VHPT-F. Of the 104 participants who completed the VPWT (39 males; mean age = 40.12 years, age range = 19–76 years), 81.7% were Caucasian, 8.7% were Asian, 5.8% were Hispanic/Latino, 2.9% were African-American, and 1.0% were Native American.

Stimuli

The VPWT used 500 images taken from the internet of forward-facing, unfamiliar, Caucasian faces from 400 identities (200 male, 200 female) that differed in lighting. Since the target identity required two images of the same identity, each trial needed five images total of four identities (four unique identities plus one target identity with two images). Using Adobe Photoshop, the faces were converted to grayscale and cropped to exclude the area from the ears outward. On trials in which the target part included the top portion of the face, the entire background, including external facial features, was removed to prevent use of non-internal face information. On part trials, non-target face portions were phase-scrambled. Each face was assigned to a part-size condition: top two-thirds, bottom two-thirds, top one-third, bottom one-third, top half, bottom half, eyes, nose, and mouth (Fig. 1). The part was then combined with a complementing face portion (either real or a phase-scrambled) to create a complete face. The target part was outlined in red (1.5-pt thick) to indicate the target part. There was approximately an equal number of trials for each face size, part size, and whole versus part conditions. The same complementing face portion was used in the study face and all three test faces (both target and distractors), such that it was not diagnostic. This method of combining face parts is used in the VHPT, a version of the composite task designed to measure individual differences that generally reveals large effect sizes for holistic processing (Richler et al., 2014). In addition, large holistic processing effects have been obtained with similar composite faces made of two different face halves (Richler & Gauthier, 2014). Different images of the target person were presented at study and test to prevent image matching.

Fig. 1 Example target parts used in the Vanderbilt Part Whole Test (VPWT). When the target part included hair, it was cropped in the study face. The top row shows top two-thirds, half and third from left to right; middle row shows bottom two-thirds, half and third from left to right; bottom row shows eyes, nose, and mouth from left to right

Procedure

The test began with instructions and practice trials (pilot 1: two cartoon, one famous face; pilot 2: two cartoon, five famous faces), followed by 81 (pilot 1) or 100 (pilot 2) experimental trials. Participants were instructed to "just try to memorize the red highlighted part" of each face. This is different from the original part-whole task, in which participants are instructed to "pay attention to the entire target face." Therefore the new task requires selective attention at encoding, a choice we made for two reasons. First, we wanted to ensure that we measured a default holistic advantage in

encoding parts in the context of a whole face (i.e., that people cannot help but encode faces holistically, rather than doing so only when the task encourages them to do so). Second, the composite task requires selective attention at encoding and produces large holistic effects on average (Richler & Gauthier, 2014) and reliable individual differences (Richler et al., 2014), so we expected that the selective attention instructions would yield similar holistic effects in the present task (at least if the two paradigms tap into the same underlying mechanism). In addition, pilot testing in a version in which the red box was not present suggested that instructing participants to memorize the entire face, rather than asking them to selectively encode a part, did not affect the results.[2] On the present version, on each trial, a study face was shown for 2 s (Fig. 2). Next, participants made an un-speeded three-alternative forced choice about which option contained the target face part that matched the identity of the study part. Response selection was not speeded to limit response bias (Richler, Mack, Gauthier, & Palmeri, 2009). Participants were given feedback during all practice trials and the first three experimental trials. Trials were blocked by target part, and presented in order of decreasing target part size (top two-thirds, bottom two-thirds, top half, bottom half, top third, bottom third, eyes, nose, mouth). Part, whole, male, and female face trials were randomized. Each pilot test and the final version took approximately 15 min to complete.

The VPWT 1.0, had 100 trials and was created based on iterative item analysis of pilot results with a larger sets of trials. As part of this process, we selected trials with a range of difficulty and on the basis of their correlation with their own condition average, and matched the difficulty of target face parts used in the whole and part conditions.

Modifications from original task

To improve the test's reliability, we used a 3-AFC format that reduces the guessing rate. In addition, unlike the original part-whole task that only used the eyes, nose, and mouth as target parts, in the VPWT, we varied the size of the target parts (Fig. 1) to create trials that vary in the extent to which they could benefit from the presence of the non-diagnostic rest of the face. Face size also varied (.59, 1.01, or 1.59 in. in face width). Size modulates holistic processing (McKone, 2009; Ross & Gauthier, 2015), so in principle varying size should increase the ability to discriminate between participants along a broader range of abilities

---

[2] On this pilot test, participants were instructed to "just try to memorize the entire face" and a black horizontal line separated the face parts (as opposed to a red box around the target face part). Other than these changes, this pilot VPWT version did not differ from other versions in any way. There was no whole-advantage on the test $t(46) = -0.98$; $p = .331$, $d = -.29$. A three-way analysis of variance of instruction set, part or whole condition, and size of target face part revealed a non-significant 3-way interaction ($F(8,57) = 0.514$, $p = .847$, $\eta_p^2 = .004$). Moreover, as found in all other VPWT versions, part and whole trials were highly correlated ($r_{50} = .53$, $p < .001$, $r_{corr} = 1.00$).

(Richler et al., 2014). Importantly, these aspects of the task are not factors of interest, but were varied to help provide discriminating information along the whole continuum of holistic processing. This is analogous to using questions with a broad range of difficulty in an intelligence test. If all questions were of equal difficulty, the test would only differentiate between those who can correctly answer questions at that difficulty level and those who cannot. By using questions from a range of difficulty levels, the test is better able to discriminate between individuals at all levels of ability. Similarly, we did not intend for face and part size factors to be analyzed and interpreted in our test (see Richler et al., 2014, for a similar strategy in the modification of the composite task).

To improve the validity of the VPWT, we added phase-scrambled information to the part condition such that part trials were presented with a phase-scrambled face in the complementing portion (see Fig. 2). We added this phase-scrambled information to keep the spatial frequency properties of part and whole trials similar (in Study 2 we verify that it had no other confounding effects). In addition, the target study and response images were different images of the same identity, so that image matching was not possible.

In the original part-whole task, participants study a whole face and then are tested on recognition of either a face part presented in isolation or within a non-diagnostic face. However, a previous study reported large study-test congruency effects with the part-whole paradigm, such that when participants studied a part, they performed better in the part versus whole condition (Leder & Carbon, 2005). Here, we matched study and test format to exclude the possibility of a context-dependent advantage, which is not the construct of interest. In this way, we aimed to ensure that our test measures how well an individual's memory of a face part is improved when that part is presented within the context of a complete face, rather than how sensitive an individual is to study-test congruency.

Because we matched study and test format, participants know as soon as they see the study face which part is relevant for the entire trial (i.e., there is no study-test incompatibility). Moreover, the target face part is outlined in red, ensuring that participants know which part will be tested as soon as they see the study face. If encoding whole faces provides an advantage and holistic encoding is under top-down control, this would encourage participants to process study faces holistically when possible. If holistic encoding is not under top-down control, it will occur automatically when whole faces, but not parts, are studied.

Because we matched the study and test format on our test, we did not expect the group level whole advantage to be large (as shown in several experiments by Leder & Carbon, 2005, when there is no contribution of study-test format incompatibility). Because we were interested in how individuals perform relative to others, the test's validity would not be threatened if we did not find that participants were on average more

Fig. 2 Example Vanderbilt Part Whole Test (VPWT) trials. Participants saw a study display (top) for 2 s, followed by a test display (bottom). Whole trials (left) showed complete faces whereas part trials (right) used phase-scrambled face parts for non-target areas. Correct responses are indicated by asterisks. Faces varied in size for both kinds of trials

accurate on whole versus part trials, as we are concerned here with the variability between individuals in their ability to use a whole face when possible.

Results

In the part-whole paradigm, variation on part trials is assumed to reflect general visual perception and face part processing abilities, whereas variation on whole trials reflects those processes as well as an additional holistic process (DeGutis, Wilmer, et al., 2013). Accordingly, holistic processing in the VPWT is operationalized by the variability on whole trials with variability on part trials regressed out (see DeGutis, Wilmer, et al., 2013). We used Guttman's λ_2 instead of Cronbach's α to calculate internal consistency of this holistic processing index because the VPWT has multiple conditions and calculation of Guttman's λ_2 incorporates the covariance between items (Guttman, 1945). Guttman's λ_2 (based on the formula in Malgady & Colon-Malgady, 1991) was .47 and .43 in Pilot 1 and Pilot 2, respectively.

Accuracy was not significantly greater on whole than part trials (part: $M = 56.6\%$, $SD = 10.7\%$; whole: $M = 57.6\%$, $SD = 9.9\%$; $t(206) = 1.03$; $p = .307$, $d = .10$), indicating no whole-advantage at the group level. Although whole and part trials had moderate internal consistency (whole $\alpha = .59$, part $\alpha = .64$), reliability of the holistic processing index (variance in whole trials after regressing out variance in part trials) was only .16 (λ_2). The holistic processing index had much lower internal consistency than either of the two conditions because

the correlation between the part and whole conditions was essentially as high as their respective reliabilities ($r_{104} = .60$, $p < .001$, $r_{corr} = .98$, $R^2 = .36$).

One concern is that trials with large target face parts (halves and two-thirds) may be responsible for the strong correlation between part and whole conditions, because larger face parts more closely resemble the whole condition. To see if this was the case, trials were grouped by target part size (small parts: eyes, nose, and mouth; medium parts: halves and thirds; large parts: two-thirds). The correlations between part and whole conditions were $r_{104} = .44$ (95% confidence interval (CI) [.27–.58], $p < .001$, $r_{corr} = 1.11$), $r_{104} = .40$ (95% CI [.22–.55], $p < .001$, $r_{corr} = 0.99$), and $r_{104} = .35$ (95% CI [.17–.51], $p < .001$, $r_{corr} = 1.15$) for large, medium, and small parts, respectively. Thus, regardless of part size, performance in the part condition almost fully accounts for performance in the whole condition when measurement error is considered. While the shared variance is numerically smaller for the smaller parts, this condition was the least reliable (Cronbach's α used here for individual conditions, small part $\alpha = .38$, whole $\alpha = .24$; medium part $\alpha = .41$, whole $\alpha = .40$; large part $\alpha = .36$, whole $\alpha = .44$).

Discussion

We attempted to create a reliable version of a modified part-whole paradigm. Although the test did not achieve sufficient reliability, the reason for this failure is interesting. The part and whole conditions are each fairly reliable independently, but

are as correlated as possible given their respective reliabilities, yielding a disattenuated correlation of $r_{corr} = .98$ (Wetcher-Hendricks, 2006). Thus, apart from measurement error, there is a near perfect correlation between the two conditions. One limitation of the VPWT 1.0 is that while the difficulty of specific parts was matched across the two conditions, the parts used were different to limit part repetition. However, versions of the original part-whole task in which the same parts were used in both conditions have produced very similar results, with part and whole conditions that were moderately reliable but lower reliability for the holistic regression index (DeGutis, Wilmer, et al., 2013). Nonetheless, the use of different parts in the two conditions also confounds any interpretation of the (absent) group-level whole advantage.

Our results suggest that performance on whole trials can essentially be perfectly predicted by performance with parts. This is very different from results in the composite paradigm, where the shared variance between critical conditions is only about 6% (Richler & Gauthier, 2014). Of course it is possible that one of the ways in which we modified the paradigm challenged its validity. To investigate this, we replicated Study 1 without the phase-scrambled noise on part trials (Study 2), when the same parts were used in whole and part conditions to equate difficulty (Studies 3 and 4), and using a version of the original part-whole paradigm (Study 3).

Study 2

One way in which the VPWT differs from the typical part-whole paradigm is the use of phase-scrambled noise in the parts condition. Although the whole advantage is thought to arise from facilitation of part memory when the target part is presented within a whole face context, it is possible that participants were also able to process the part in the phase-scrambled noise context holistically. For example, participants could have interpreted the noise parts as a disguise, or perceived face parts in the noise itself. We tested this using a version of the VPWT in which parts were presented in isolation on a white background (i.e., without phase-scrambled noise), similar to the original task.

Methods

Participants

One hundred and ten participants were recruited from Amazon Mechanical Turk and were compensated US$0.50. Six participants were excluded from the analyses for failure to follow instructions, leaving 104 participants (34 males; mean age = 40.70 years, age range = 20–76 years), of whom 80.8% were Caucasian, 7.7% were African American, 7.7% were Hispanic/Latino, 2.9% were Asian, and 1.0% identified as other.

Vanderbilt Part Whole Test 2.0

The VPWT 2.0 was identical to the VPWT 1.0, except that phase-scrambled irrelevant face parts were removed from part trials. The test took approximately 15 min to complete.

Results

Once again accuracy was not significantly different between part and whole trials (part: $M = 56.2\%$, $SD = 8.6\%$; whole: $M = 56.0\%$, $SD = 8.7\%$; $t(206) = -0.30$; $p = .762$, $d = -0.02$), indicating no whole-advantage at the group-level. Whole and part conditions had high internal consistency (whole $\alpha = .72$, part $\alpha = .71$), but the reliability of the holistic processing index (whole trial variance with part trial variance regressed out) was very low (.07), with part and whole trials as strongly correlated as possible given the measurement error ($r_{104} = .73$, $p < .001$, $r_{corr} = 1.02$, $R^2 = .53$). The correlations between part and whole conditions were similar across target part size (large: $r_{104} = .58$, 95% CI [.44–.69], $p < .001$, $r_{corr} = 1.18$; medium: $r_{104} = .53$, 95% CI [.37–.65], $p < .001$, $r_{corr} = 1.01$; small: $r_{104} = .32$, 95% CI [.14–.48], $p < .001$, $r_{corr} = 1.31$). These results are similar to Study 1, and again the smaller parts condition was the least reliable (small part $\alpha = .14$, whole $\alpha = .42$; medium part $\alpha = .50$, whole $\alpha = .55$; large part $\alpha = .55$, whole $\alpha = .44$). Importantly, even with the small parts, the correlation between the part and whole conditions is as large as can be expected based on their respective reliabilities.

Discussion

The results for the VPWT 2.0 in Study 2 were highly similar to those from the VPWT 1.0 in Study 1, where parts were presented in a phase-scrambled context. Thus, presenting face parts within a phase-scrambled context is not responsible for the very high correlation between part and whole conditions.

Study 3

In two studies using the VPWT, we found evidence that part and whole trials were strongly correlated, tentatively suggesting the use of a similar processing strategy in both conditions. This conflicts with the assumption that participants engage holistic processing for whole face trials, but rely on feature processing for part trials (Tanaka & Farah, 1993). However, it is possible that some of the other modifications we made to the original part-whole task made our paradigm fundamentally different from the original paradigm (Tanaka & Farah, 1993). In addition, only small parts are used in the original

design, so with more trials it should be possible to get measurements that are more reliable in small part and whole conditions and verify that our results hold for small parts only. In Study 3 we directly compare the VPWT and the original part-whole task.

Methods

Participants

One hundred and six participants recruited from Amazon Mechanical Turk completed the original part-whole paradigm, followed by the VPWT 3.0. Participants were compensated US$2.00 for successfully completing both tasks. Eight participants were excluded for failure to follow instructions, leaving 98 participants (35 males; mean age = 38.01 years, age range = 20–66 years), of whom 78.6% were Caucasian, 7.1% were African American, 6.1% were Hispanic/Latino, 6.1% were Asian, and less than 1.0% identified as other.

VPWT 3.0

In the previous studies, we found that part and whole trials were strongly correlated and wanted to verify that this would still hold when part and whole trials were more precisely equated for difficulty by using the same target parts in both conditions (trials were chosen from version 1.0 to maximize reliability and range of difficulty). Because performance on

the target mouth condition (mouth without the chin) was at chance for most previous iterations of the task, this condition was not included in the VPWT 3.0. The final test consisted of 212 trials and took approximately 20 min to complete. None of the face stimuli in the VPWT 3.0 and original part-whole test are images of the same individuals.

Original part-whole test

Stimuli We obtained stimuli from the commonly used part-whole paradigm (DeGutis, Wilmer, et al., 2013, used with permission from James Tanaka, University of Victoria; Tanaka, Kiefer, & Bukach, 2004). The target stimuli are faces made of different eye, nose, and mouth images overlaid on a face template. The same male and female external contour face templates (seen in Fig. 3) were used on all trials. All faces were Caucasian male and female faces. Each target face was made of a completely unique set of the three face features. Foil faces were created by changing the eyes, nose, or mouth of a target face. In this way, only one feature in the foil faces differed from the target face. Because previous work reported low reliability for this task (DeGutis, Wilmer, et al., 2013, whole residual α = .19), we attempted to increase reliability by doubling the number of trials. The new stimuli were created using the same face template as the original task, but with entirely new face parts, obtained from the internet. The part trials were created by cropping the whole target or foil face so that only the target face part was visible (Fig. 3, right image).
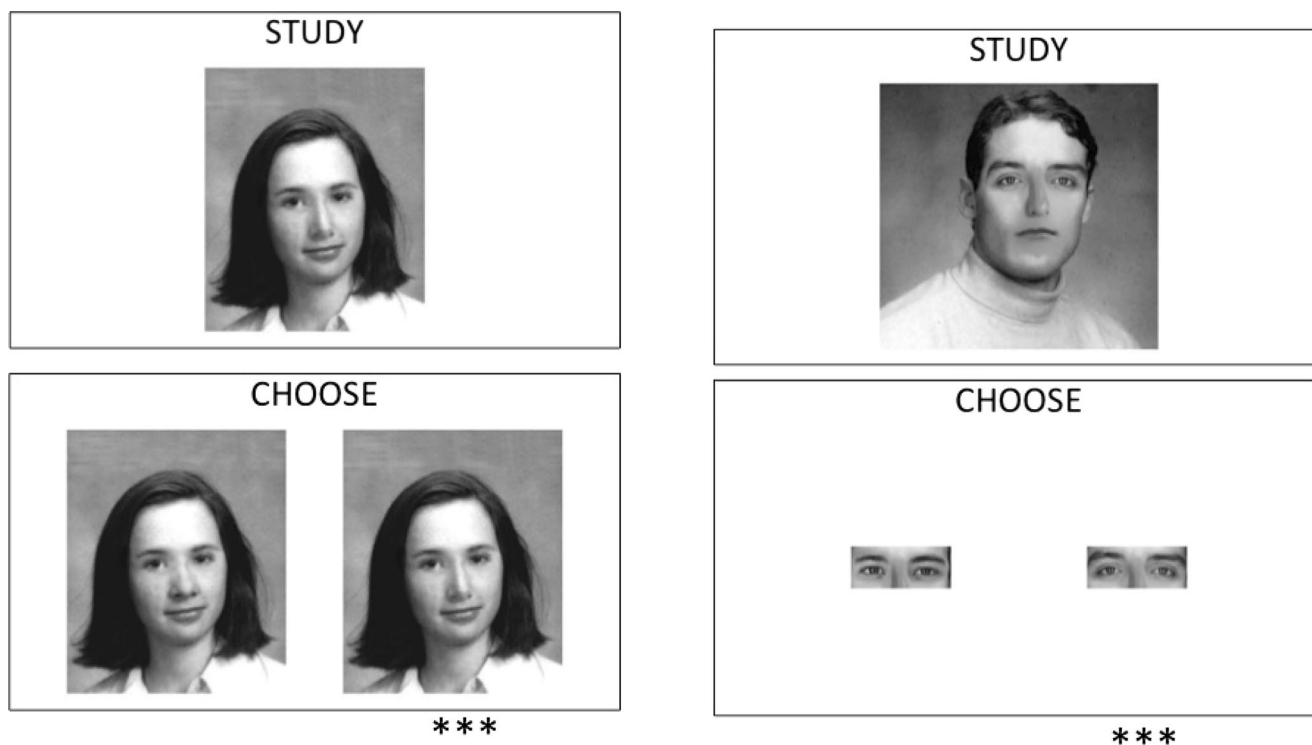
Fig. 3 Example original part-whole trials. Participants saw a study display (top) for 1 s, followed by a mask for 500 ms and then the test display (bottom). Whole trials (left) presented a complete face whereas part trials (right) presented the target face parts in isolation. Correct responses are indicated by asterisks

Each face part was repeated four times total throughout the test (once in a part trial, once as a target part in a whole trial, and twice as the non-target part in a whole trial). Foil parts were repeated twice (once in a part trial and once in a whole trial). In total, there were 144 trials, with an equal number of male and female, whole and part, and target face part (eye, nose, or mouth) trials. The instructions were identical to those used in DeGutis, Wilmer, et al., 2013.

Procedure

Participants first completed the original part-whole task followed by the VPWT 3.0 (212 trials). In the original part-whole task, each trial started with a central fixation for 500 ms followed by a whole target face for 1 s. Then, after a 500-ms mask, the target and foil face were presented side by side as either wholes or parts (see Fig. 3). Participants made a two-alternative forced choice by mouse click. There was a 250-ms inter-stimulus interval. Part and whole trials and new and old stimuli were randomized. The entire experiment took approximately 45 min (20 min for the original part-whole task, 25 min for the VPWT 3.0).

Results

The correlations across conditions in the two tests are reported in Table 1. Results from the VPWT 3.0 were similar to previous studies. Accuracy did not significantly differ between part and whole trials (part: M = 51.7%, SD = 9.2%; whole: M = 50.8%, SD = 9.4%; $t(194)$ = 1.40; p = .164, d = −0.09), indicating no whole-advantage at the group level. The strong correlation between part and whole conditions replicated (r_{98}= .79, p < .001, r_{corr} = 1.02, R^2 = .62), with acceptable internal consistency for both whole and part conditions (whole α = .77, part α = .76). Again, the reliability of the holistic processing index (whole trial variability not accounted for by part trial variability) was null (λ_2 = −.03).

Table 1 Bottom left corner of the table (**bold**) shows the Pearson Product–moment Correlations (N = 98). Upper right corner shows the same correlations, disattenuated for measurement error based on reliability of the measurements

	1.	2.	3.	4.	5.	6.
1. Original part	--	.98	.00	.83	.80	†
2. Original whole	**.70*****	--	1.63	.91	.81	†
3. Original resid	**.00**	**.71*****	--	.95	.74	†
4. VPWT part	**.58*****	**.70*****	**.41*****	--	1.04	†
5. VPWT whole	**.57*****	**.63*****	**.32****	**.79*****	--	†
6. VPWT resid	**.18**	**.12**	**-.01**	**.00**	**.61*****	--

† Disattenuated correlations with VPWT residuals are not included because reliability was too low

*p < .05, **p < .01, ***p < .001

The original part-whole paradigm showed a significant whole condition advantage at the group level (part: M = 68.8%, SD = 8.9%; whole: M = 77.5%, SD = 9.9%; $t(194)$ = −6.49; p < .001, d = 0.93). The whole and part conditions had acceptable internal consistency (whole α = .78, part α = .66), but consistent with earlier experiments, reliability was low for the holistic processing index reliability (.24), and part and whole trials were strongly correlated (r_{98}= .70, p < .001, r_{corr} = .98, R^2 = .49).

Discussion

We compared the original part-whole paradigm and the modified VPWT paradigm (VPWT 3.0). By doubling the number of trials in the original task, we achieved higher reliability of whole and part trials, (whole α = .77, part α = .76, compared with .65 and .43, respectively, in DeGutis, Wilmer, et al., 2013). The strong correlation between corresponding conditions across the original part-whole paradigm and the VPWT suggests that they measure similar constructs, despite differences in procedure and stimuli. More importantly, after accounting for measurement error, we find that whole condition performance can essentially be perfectly predicted from part condition performance. This is the case in both the VPWT and the original paradigm.

In Study 3 reliability of the holistic processing index was very low. Indeed, the only report we can find in the literature of even moderately reliable holistic processing in the part-whole paradigm is for Asian faces in Caucasian observers (λ_2 = .48, in the same paper, it was .33 for Caucasian faces, DeGutis, Mercado, et al., 2013). It is possible that when observers are not as familiar with the category of faces (e.g., from a less familiar race), whole and parts are not processed as similarly as familiar categories, but this should be confirmed in future studies. Moreover, the expectation is that other-race faces are processed *less* holistically than same-race faces (Tanaka et al., 2004), and other tasks have revealed comparable holistic processing for same- and other-race faces (Harrison, Gauthier, Hayward, & Richler, 2014; Hayward, Rhodes, & Schwaninger, 2008).

The goal of creating the VPWT was to develop a reliable holistic processing measure that we could then use to investigate how face recognition ability and holistic processing relate. Our results suggest that in both the original instantiation of the part-whole paradigm and our revised version, part and whole trials variation almost completely overlap. However, the reliabilities of part and whole conditions are not perfect, so this conclusion comes from correlations that have been adjusted based on a substantial amount of attenuation (in other words, these correlations are theoretical). Therefore, to provide converging evidence on the similarity of processing in part and whole conditions, in Study 4 we tested the extent to which performance on part and whole conditions predicts

performance on a highly reliable test of face recognition, the CFMT (Duchaine & Nakayama, 2006). In previous work (DeGutis, Mercado, et al., 2013) with the original part-whole paradigm, the correlation between the CFMT and part condition was moderate ($r = .45$, $r_{corr} = .67$), and that between the CFMT and whole condition was slightly higher ($r = .54$, $r_{corr} = .79$). In another study, DeGutis, Wilmer, et al., (2013) found similar correlations between each condition and the CFMT (part condition: $r = .44$, $r_{corr} = .65$; whole condition: $r = .63$, $r_{corr} = .80$). In these two studies, there was also a significant correlation between the CFMT and holistic processing in the whole part task using a regression index (DeGutis, Mercado, et al., 2013; $r = .47$, 95% CI [.3–.66]; DeGutis, Wilmer, et al., 2013; $r = .46$, 95% CI [.18–.67]).

In Study 4, we wanted to revisit the correlations between part and whole condition performance and the CFMT for a number of reasons. First, the correlations in the two studies by DeGutis et al. have large confidence intervals because the samples were relatively small (43 and 53 participants, respectively). Second, these studies did not estimate the face-specific nature of this relation because only correlations with a face recognition measure were tested. Third, in the original part-whole paradigm (used in Study 3), participants always study a whole face. Because Leder and Carbon have shown that this procedure produces a sizeable study–test compatibility effect (Leder & Carbon, 2005), variability in this effect may drive the correlation between the original part-whole task and the CFMT. In contrast, study and test conditions are matched in the VPWT, so correlations cannot be driven by variability in the study-test compatibility effect.

Study 4

Here we assessed how performance on part and whole trials in the VPWT relate to extant measures of face and object recognition. The CFMT (Duchaine & Nakayama, 2006) is a widely used and highly reliable (~.8) measure of face recognition ability. The CFMT shows low to moderate correlations with tests of object recognition that use a similar task format (Dennett et al., 2012; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Van Gulick, McGugin, & Gauthier, 2015), and the ability measured is highly heritable and independent from general cognitive ability and intelligence (Richler et al., General object recognition is specific: Evidence from novel and familiar object, manuscript in preparation; Shakeshaft & Plomin, 2015; Wilmer et al., 2010).

More relevant to our current goals, the CFMT has also been used to investigate whether face recognition ability relates to holistic processing. This test was originally created to measure the "special mechanism used to recognize upright faces," assumed to depend on holistic or configural representations (Duchaine & Nakayama, 2006). However, the only evidence

that the CFMT taps into such representations is that participants perform better on the test when face stimuli are upright rather than inverted. This represents indirect evidence at best, given that strong inversion effects are also obtained for non-face objects that are not thought to be processed holistically (e.g., Ashworth, Vuong, Rossion, & Tarr, 2008). Past work that found a small but significant correlation between the original part-whole paradigm and the CFMT did not include a measure of non-face object recognition ability (DeGutis, Mercado, et al., 2013; DeGutis, Wilmer, et al., 2013). The untested assumption is that the additional process engaged by the whole but not part condition is one that is useful for faces specifically. To test this assumption, we used the Vanderbilt Expertise Test (VET) as an object recognition measure that is similar in format to the CFMT (see McGugin et al., 2012).

Because we found that part and whole conditions strongly correlate in Studies 1–3, we expect that these two conditions will relate to a similar extent to the CFMT and VET. We performed a hierarchical regression analysis to quantify any variance in the whole condition that might predict face recognition ability specifically, beyond what is predicted by the part condition.

Methods

Participants

Two hundred fifty participants were recruited from Amazon Mechanical Turk to complete the CFMT and VET-motorcycle. We contacted participants one day after they completed the CFMT and VET-motorcycle and offered them the opportunity to complete the Vanderbilt Face Matching Test (not reported here), VETs with the other four categories, and VPWT. Participants were compensated US$0.70 for completing the CFMT and VET-motorcycle, and a total of US$5.00 for the other tasks. We purposefully recruited a large number of participants so that the subset that chose to complete all tasks would have sufficient power. 115 participants chose to complete the four additional tasks. Eleven participants were excluded for failure to follow instructions or failure to respond correctly to both VET catch trials. Of the 104 participants (35 males; mean age = 37.60 years, age range = 20–70 years) who satisfactorily completed all tasks, 85.6% were Caucasian, 4.8% were Asian, 3.9% were Hispanic/Latino, 2.9% were African-American, and 2.8% identified as other.

Cambridge Face Memory Test (CFMT)

We used the long version of the CFMT (Russell, Duchaine, & Nakayama, 2009). Participants studied six Caucasian gray-scale male target faces, then on each trial had to correctly identify the target face presented with two foil faces. The first block of 18 trials showed target faces in the studied viewpoint.

The second block of 30 trials required participants to identify the target across variations in lighting and viewpoint, and in the third block of 24 trials Gaussian noise was added to novel target images. The last block of 30 trials was the most difficult, with uncropped targets and target faces in profile, both with additional noise added. Participants were allowed to study the target images between each block and responses were unspeeded. The CFMT takes approximately 10 min to complete.

Vanderbilt Expertise Test (VET)

Participants studied six exemplars from an object category for 20 s, and were then tested with identical exemplars for six trials with feedback. This was followed by another 20-s study period, then six more trials with feedback. Finally, participants completed 36 trials where the target exemplar was not an identical image to the study exemplar and no feedback was provided. Different versions of the VET with different categories have been used (e.g., McGugin et al., 2012; Van Gulick et al., 2015). Here, we used VETs for five categories: motorcycles, planes, birds, houses, and butterflies. To produce a domain-general estimate of performance with objects, we used an average of all five categories. The five VETs were positively correlated with each other, inter-item correlation range = .33–.56, all ps < .05, Cronbach's α for the whole test = .93. Each VET for a single domain takes approximately 10 min to complete.

Results

Mean accuracies and reliabilities are reported in Table 2. Removing non-Caucasian participants did not significantly change mean accuracy for any test ($t(176) < 0.42$, $p > 0.674$).

We calculated Pearson Product–moment correlations between the VET (average accuracy across all five categories), CFMT, VPWT 3.0 whole residuals (performance on whole trials, regressing out performance on part trials), VPWT 3.0 whole trial accuracy, and VPWT 3.0 part trial accuracy. Correlations and correlations disattenuated for measurement error are shown in Table 3.

Table 2 Summary statistics for the tests used in Study 4 ($N = 104$)

	Mean	SD	Reliability
CFMT	61.4%	11.3%	.868[a]
VET combined	65.6%	6.0%	.929[a]
VPWT Whole Acc	55.9%	8.4%	.703[a]
VPWT Part Acc	53.0%	8.1%	.675[a]
VPWT Whole Residuals	0.03%	6.0%	.095[b]

[a] Cronbach α

[b] Residual Reliability, λ_2

Table 3 Bottom left corner of the table (**bold**) shows the Pearson Product–moment Correlations (N = 104). Upper right corner shows the same correlations, disattenuated for measurement error based on reliability of the measurements (no significance values are assigned as significance tests are inappropriate for corrected correlations)

	1.	2.	4.	5.	6.
1. CFMT	--	.46	.71	.68	†
2. VET	.41***	--	.60	.65	†
4. Whole Acc	.55***	.48***	--	1.02	†
5. Part Acc	.52***	.51***	.70***	--	†
6. VPWT whole residuals	.26**	.17	.71***	.00	--

† Disattenuated correlations with VPWT residuals are not included because reliability was too low

*p < .05, **p < .01, ***p < .001

Once again, the internal consistency of the VPWT 3.0 holistic processing measurements (whole residuals) was extremely low (.02). The internal consistencies were much higher for part (α = .68) and whole (α = .69) conditions individually and, replicating prior results, performance in the two conditions was strongly correlated (r_{138} = .72, p < .0001, R^2 = .52). In fact, the disattenuated correlation again suggests that the shared variance between part and whole conditions essentially accounts for all the non-error variance (Table 3). Accuracy was higher on whole than part trials (Table 2, $t(206)$ = 4.67, p < .0001, d = 0.35), indicating a whole-advantage at the group level.

Because the VPWT 3.0 reliability for holistic processing was extremely low, the disattenuated correlations have very low precision. For instance, the disattenuated correlation between the VPWT 3.0 whole residuals and CFMT is r_{104} = .89, but the 95% CI is very large [.23–1.50] (even though the CI extends beyond a possible maximal correlation of 1.0, it nonetheless provides information about uncertainty of the estimate).

To compare our results to previous work in which CFMT and holistic processing in the standard whole part task were correlated (DeGutis, Mercado, et al., 2013; DeGutis, Wilmer, et al., 2013), we performed a two-step hierarchical regression (Table 4). Part accuracy accounted for 26.2% of CFMT

Table 4 Hierarchical regression predicting Cambridge Face Memory Test (CFMT) performance ($N = 104$). Part accuracies (Acc) are entered as a first step, followed by whole Acc. ΔR^2 column shows the change in variance accounted for in decimal form

Variable	β	SE(β)	ΔR^2
Step 1			
Part Acc	.72	.12	.26***
Step 2			
Whole Acc	.25	.11	.04*

*p < .05, **p < .01, ***p < .001

variance, and whole accuracy accounted for an additional 3.8%, which was a small but significant increase. Thus, as in previous work, there is a small amount of variance in CFMT performance that whole accuracy does predict beyond part accuracy.

What has not been addressed before is how specific this whole effect is to face processing. To this end, we performed a second hierarchical regression on CFMT, adding VET accuracy in step 1 to preserve only face-specific ability before entering part and whole accuracy as predictors (Table 5). VET accuracy accounted for 16.2% of CFMT variance, consistent with prior work (Gauthier et al., 2014; Van Gulick et al., 2015). Part accuracy accounted for 10.5% of face-specific variance and, critically, whole trial accuracy did not account for a significant amount of face-specific variance.

Discussion

Using the VPWT, we explored how part and whole trials relate to face and object recognition measures. As in Studies 1–3, we found that part and whole trials were as correlated with each other as was possible given measurement error. The VPWT 3.0 index of holistic processing was weakly but significantly correlated with the CFMT, replicating prior work. However, the relation does not account for face-specific variance in the CFMT, as the effect did not survive when object recognition ability was regressed out. Performance with the whole was a little better at predicting face recognition ability than performance with parts but this was not a face-specific effect. This effect could be due to strategies that participants apply to whole objects more generally, and may reflect more general global processing abilities (Milne & Szczerbinski, 2009).

General discussion

We modified the classic part-whole paradigm to create a measure of the whole advantage that would be more reliable than

Table 5 Hierarchical regression predicting Cambridge Face Memory Test (CFMT) performance ($N = 104$). Vanderbilt Expertise Test (VET) average (Avg) accuracies are entered as a first step, followed by part accuracies (Acc) and then whole Acc. ΔR^2 column shows the change in variance accounted for in decimal form

Variable	β	$SE(\beta)$	ΔR^2
Step 1			
VET_Avg	.48	.11	.16***
Step 2			
Part Acc	.43	.12	.11***
Step 3			
Whole Acc	.18	.11	.02

*$p < .05$, **$p < .01$, ***$p < .001$

the measure used in prior work. Surprisingly, once the part and whole conditions were matched for difficulty and achieved high reliability, we found little evidence for variability in the whole trials unaccounted for by variability in the part trials despite the fact that on average participants often performed better with whole faces than parts.

Admittedly, the conclusions from Studies 1–3, that part and whole processing shared all the variance we could reliability measure, seems challenged by Study 4, in which whole accuracy accounted for a little more variance in face recognition ability than part accuracy. Given the small correlation between face recognition and holistic processing in Study 4 and the large confidence intervals we would have to place around disattenuated correlations between part and whole conditions in earlier studies, these discrepant results may be a reflection of the limitations of measurements in this field. Accordingly, an important goal for future research in this domain is to continue efforts to improve the psychometric properties of our tasks.

Although many studies have used the part-whole paradigm, fewer have approached the paradigm from an individual differences perspective. Most of our discussion has focused on comparisons to the individual differences work by DeGutis, Mercado, et al. (2013), DeGutis, Wilmer, et al. (2013) who used the original version of the part-whole task. In Study 3, we found evidence of convergent validity for the VPWT, since the whole and part trials from the VPWT correlated with the whole and part trials from the original part-whole task respectively. This is informative, suggesting that the ability measured in that task is robust across a range of stimuli and instructions. Study 4 also qualitatively replicated DeGutis et al., with the whole-advantage correlating with CFMT performance. However, our effect was less than 4%, considerably less than the ~22% obtained by DeGutis, Mercado, et al. (2013), DeGutis, Wilmer, et al. (2013). It is possible that their effect size was inflated due to smaller sample sizes (Halsey, Curran-Everett, Vowler, & Drummond, 2015), the repetition of faces in both tasks (Richler et al., 2015), and/or the contribution of study-test compatibility effects (Leder & Carbon, 2005).

We also extended this result and tested the validity of the construct measured in the whole-part paradigm by also measuring object recognition ability. Object recognition ability is typically not measured in studies concerned with holistic processing of faces, likely due to the assumption that holistic processing is simply not relevant to non-face object recognition. However, there is about 20% shared variance between CFMT and VET performance, and our results illustrate that controlling for object recognition ability is an important part of testing theories about holistic processing.

How do we reconcile these results with the larger literature that uses the whole-advantage in the part-whole paradigm as evidence of holistic processing in group studies? First, the

patterns of correlations we observed were the same regardless of whether or not there was a whole-advantage at the group level. The average effect was smaller in the VPWT than it generally is in the standard task, which we expected based on many demonstrations that the bulk of the whole advantage is due to study-test compatibility effect (i.e., a part advantage when parts are studied; Leder & Carbon, 2005). Our results should not be taken to show that there is no holistic processing of whole faces. Measures using variations on the composite task (like the VHPT-F) that operationalize holistic processing as a failure of selective attention show strong evidence for reliable and stable individual differences in holistic processing (Richler et al., 2014). Because the encoding of face parts in the whole condition of the VPWT is essentially the same (same instructions and stimuli) as in the VHPT-F, we can infer that in some way that can be detected by the congruency effect, these faces could be processed holistically and to a variable degree by different subjects. Therefore, our specific conclusions focus on the operationalization of holistic processing in the various versions of the part-whole paradigm that we tested here, suggesting that it does not capture a substantial amount of variability in holistic processing, in which case it may be of little use in studies of individual differences. Future work may explore if such variability can be obtained in other versions of the part-whole paradigm, or assess whether the current results with the original conditions used in Study 3 are stable in exact replications. Importantly, the present work offers a clear test of whether any version of the part-whole paradigm can be deemed useful to study individual differences in holistic processing: that there should be considerable variance in the whole condition that is not accounted for by a part condition.

References

Ashworth, A. R., III, Vuong, Q. C., Rossion, B., & Tarr, M. J. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition, 16*(6), 754–784.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia, 49*(9), 2541–2552.

Boggan, A. L., Bartlett, J. C., & Krawczyk, D. C. (2012). Chess masters show a hallmark of face processing with chess. *Journal of Experimental Psychology: General, 141,* 37–42.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*(1), 605–634.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5.

Bukach, C. M., Philips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception & Psychophysics, 72,* 1865–1874.

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review, 17*(2), 155–160.

Carbon, C. C., Grüter, T., Weber, J. E., & Lueschow, A. (2007). Faces as objects of non- expertise: Processing of thatcherised faces in congenital prosopagnosia. *Perception, 36*(11), 1635–1645.

Cho, S. J., Wilmer, J., Herzmann, G., McGugin, R. W., Fiset, D., Van Gulick, A. E., … Gauthier, I. (2015). Item response theory analyses of the Cambridge Face Memory Test (CFMT). *Psychological Assessment, 27*(2), 552.

Chua, K. W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts.

DeGutis, J., Mercado, R. J., Wilmer, J., & Rosenblatt, A. (2013). Individual differences in holistic processing predict the own-race advantage in recognition memory. *PLoS ONE, 8*(4), e58253.

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. doi:10. 1016/j.cognition.2012.09.004

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.

Donnelly, N., Cornes, K., & Menneer, T. (2012). An examination of the processing capacity of features in the Thatcher illusion. *Attention, Perception & Psychophysics, 74*(7), 1475–1487.

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.

Fitousi, D. (2015). Composite faces are not processed holistically: Evidence from the Garner and redundant target paradigms. *Attention, Perception & Psychophysics, 77*(6), 2037–2060.

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Van Gulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. Journal of Vision, 14(8), 7–7.

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. doi:10.1016/ S0042-6989(97)00442-2

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*(4), 255–282.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods, 12*(3), 179–185.

Harrison, S. A., Gauthier, I., Hayward, W. G., & Richler, J. J. (2014). Other-race effects manifest in overall performance, not qualitative processing style. *Visual Cognition, 22*(6), 843–864.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do participant pool participants. *Behavior Research Methods, 48*(1), 400–407.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition, 106*(2), 1017–1027.

Leder, H., & Carbon, C. C. (2005). When context hinders! Learn–test compatibility in face recognition. *The Quarterly Journal of Experimental Psychology. A, 58*(2), 235–250.

Malgady, R., & Colon-Malgady, G. (1991). Comparing the reliability of difference scores and residuals in analysis of covariance. *Educational and Psychological Measurement, 51*(4), 803–807. doi:10.1177/001316449105100401

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1–23.

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

Maurer, D., & Young, R. E. (1983). Newborn's following of natural and distorted arrangements of facial features. *Infant Behavior & Development, 6*, 127–131. doi:10.1016/S0163-6383(83)80018-6

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*(C), 10–22. doi:10.1016/j.visres.2012.07.014

McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research, 49*(2), 268–283.

Milne, E., & Szczerbinski, M. (2009). Global and local perceptual style, field- independence, and central coherence: An attempt at concept validation. *Advances in Cognitive Psychology, 5*, 1–26.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical Turk. *Judgment and Decision making, 5*(5), 411–419.

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), 1–14. doi: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*, 1–12.

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., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing happens at a glance. *Vision Research, 49*, 2856–2861.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00553/abstract

Ross, D. A., & Gauthier, I. (2015). Holistic processing in the composite task depends on face size. *Visual Cognition, 23*(5), 533–545.

Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite task measurements of holistic face processing. *Behavior Research Methods, 47*, 736–743.

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. doi:10.3758/PBR.16.2.252

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences, 112*(41), 12887–12892.

Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*(2), 225–245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1–B9. doi:10.1016/j.cognition.2003.09.011

Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2015). Measuring non-visual knowledge about object categories: The semantic vanderbilt expertise test. *Behavioral Research Methods*.

Wetcher-Hendricks, D. (2006). Adjustments to the correction for attenuation. *Psychological Methods, 11*(2), 207. doi:10.1037/1082-989X.11.2.207

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*(11), 5238–5241. doi:10.1073/pnas.0913053107

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*(9), 1108–1117.