



The face inversion effect does not provide a pure measure of holistic face processing

Christian Gerlach¹ · Erik Mogensen¹

Accepted: 13 December 2022 / Published online: 9 January 2023
© The Psychonomic Society, Inc. 2023

Abstract

It is widely held that upright faces are processed more holistically than inverted faces and that this difference is reflected in the face inversion effect. It is not clear, however, how the inversion effect can best be measured, whether it is task specific, or even whether it specifically correlates with processing of upright faces. We examined these questions in a large sample ($N = 420$) who provided data on processing of upright and inverted stimuli in two different tasks with faces and one with objects. We find that the inversion effects are task dependent, and that they do not correlate better among face processing tasks than they do across face and object processing tasks. These findings were obtained regardless of whether inversion effects were measured by means of difference scores or regression. In comparison, only inversion effects based on regression predicted performance with upright faces in tasks other than those the inversion effects were derived from. Critically, however, inversion effects based on regression also predicted performance with inverted faces to a similar degree as they predicted performance with upright faces. Consequently, and contrary to what is commonly assumed, inversion effects do not seem to capture effects specific to holistic processing of upright faces. While the present findings do not bring us closer to an understanding of which changes in cognitive processing are induced by inversion, they do suggest that inversion effects do not reflect a unitary construct; an implicit assumption that seems to characterize much of the research regarding face processing.

Keywords Difference scores · Face inversion effect · Face processing · Holistic processing · Individual differences · Regression

Introduction

It is widely assumed that faces are processed more holistically—as perceptual wholes or integrals—than non-face objects such as houses, cars, or furniture (McKone & Robbins, 2011; Rossion, 2015), which in return are presumed to be processed in a more parts-based manner (Farah et al., 1998; Richler & Gauthier, 2014). Despite this, there is no commonly accepted definition of what ‘holistic’ processing is (Piepers & Robbins, 2012; Richler et al., 2012), and perhaps because of this some have argued that it is critical to ground the construct ‘holistic’ in experimental paradigms that can operationally define it. Three such paradigms are considered central (Rezlescu et al., 2017; Tanaka & Gordon, 2011): (i) the composite face paradigm (Young et al., 1987), the part–whole paradigm (Tanaka & Farah, 1993; Tanaka & Simonyi, 2016), and the face inversion paradigm (Yin, 1969).

In recent years, there have been several attempts to examine how the holistic effects measured by these paradigms relate to individual differences in face recognition performance. The results so far have been inconsistent. A few studies have found that (some) of the measures can account for some of the variance in face recognition (DeGutis et al., 2013; Rezlescu et al., 2017; Richler et al., 2011; Wang et al., 2012) but others have not (Konar et al., 2010; Rezlescu et al., 2017; Richler et al., 2015; Richler & Gauthier, 2014; Verhallen et al., 2017). To make matters worse, it has also been found that the three measures of holistic processing do not correlate to any great extent—if at all—which raises concern as to whether they tap the same construct, as is commonly assumed (for further discussion of this see Rezlescu et al., 2017). Such a concern has even been raised for one of the paradigms considered in isolation—the composite face paradigm—which may yield different results depending on which particular design of the paradigm (partial or complete) is applied (Richler & Gauthier, 2014; Rossion, 2013). Recently, we have been confronted with

✉ Christian Gerlach
cgerlach@health.sdu.dk

¹ Department of Psychology, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark

a similar concern regarding the face inversion paradigm, which is the paradigm of focus here.

The face inversion effect refers to the finding that processing of faces becomes disproportionately more disrupted by inversion than processing of non-face objects (Yin, 1969), a finding that has been replicated in a great number of studies (Bruyer, 2011) and examined in several (sub)disciplines including developmental psychology, cognitive psychology, neuropsychology, experimental psychology, comparative psychology and neuroscience (Cashon & Holt, 2015; Griffin, 2020; James et al., 2013; Klargaard et al., 2018; Leder et al., 2017; Rhodes et al., 1993). While the face inversion paradigm does not assess holistic processing directly—because it does not involve manipulations of holistic processes per se (McKone & Robbins, 2011; Tanaka & Simonyi, 2016)—the face inversion effect is usually taken to reflect that upright faces are processed holistically whereas a parts-based analysis is the only available processing option for inverted faces (Piepers & Robbins, 2012; Rossion, 2008). Indeed, inversion abolishes the holistic effects typically found with the composite face paradigm (Rossion & Boremanse, 2008) and the part-whole paradigm (Tanaka & Sengco, 1997).

If the face inversion effect does reflect holistic processing—a proposition that will be challenged in the Discussion—one might suspect this effect to be reduced in individuals with face processing deficits, provided of course that their deficits are related to holistic processing. In partial support¹ of this proposition we recently found that the face inversion effect was reduced in a group of individuals with developmental prosopagnosia (Klargaard et al., 2018); a syndrome characterized by lifelong impairment in face recognition in the absence of brain damage. This was found across two widely used tasks in neuropsychological studies of face processing: The Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006) and the Cambridge Face Perception Test (CFPT) (Duchaine, Germine, et al., 2007a). The interesting part of this study in the present context was that the magnitude of the inversion effects measured by the two tasks did not correlate, neither for the individuals with developmental prosopagnosia, the control participants or for the two groups combined. This might suggest that the ‘face inversion effect’—just like the composite face effect—does not reflect a singular phenomenon. It is also

possible, however, that the lack of correlation between the two inversion effects was caused by limited statistical power given that the sample consisted of only 16 individuals with developmental prosopagnosia and 32 controls. In addition, the reliability of the CFPT inversion measure was found to be rather unsatisfactory (r_{sb} ranged from $-.09$ for the control sample to $.19$ for the sample with developmental prosopagnosia). This is problematic because the reliabilities of two measures limits the magnitude of the correlation that can be observed between them (Spearman, 1904). It is also worth noting that our measures of inversion effects were based on difference scores (accuracy for upright faces—accuracy for inverted faces) which are considered problematic for two reasons: (i) difference scores are often less reliable than their component scores (Peter et al., 1993), and (ii) difference scores may be confounded in the sense that both the condition of interest (here upright faces) and the control task (here inverted faces) may correlate with the difference score. If this is so, one could argue that the difference score is an unspecific measure. In the present case it would mean that the effects measured not only reflect holistic processing in the upright condition but also processes engaged by the inverted condition.

The general point regarding the unspecific nature of difference scores has been argued strongly by DeGutis et al. (2013) precisely with regards to the role of holistic processing in face recognition. They suggest that a better solution than to subtract scores is to regress the effect of the control condition (inverted faces) from the condition of interest (upright faces). The residuals left by regression should then yield a more specific measure of that part of the performance with upright faces that cannot be accounted for by performance with inverted faces, that is, the effect of holistic processing. By using such a regression approach, DeGutis et al. (2013) were able to establish that two measures of holistic processing—the part-whole and the composite face effect—both correlated with an independent measure of face recognition and with each other. This was not revealed by a similar analysis based on difference scores. In addition, the measures based on regression proved to have higher reliability than the measures based on subtraction (for similar results see Rezlescu et al., 2017).

Even though the regression approach seems attractive, it is not without concerns. Basically, it constitutes a departure from an individual difference approach in that the residual of a given participant reflects not only the individual’s performance but also the performance of the rest of group on which the regression model is based. Hence, the regression approach is subject to sampling error and this aspect should be considered in the context of the suggestion that stable estimates of correlations (based on Pearson’s r) may require sample sizes of $N = 250$ (Schönbrodt & Perugini, 2013); a sample size considerably beyond that used in most studies of holistic processing (for exceptions see Boutet

¹ The support was partial because the individuals with developmental prosopagnosia not only performed poorly with upright faces—which according to the theory are subject to holistic processing—but also with inverted faces. Similarly, one might expect individuals with exceptional face processing skills to show increased inversion effects. While there is some empirical support for this hypothesis these studies were based on small samples ($N = 1–6$) (Bennetts et al., 2017; Bobak et al., 2016; Russell et al., 2009). In comparison, in a larger sample with 14 ‘super-recognizers’ no increased face inversion effect was found using the CFMT and the CFPT (Hendel et al., 2019).

et al., 2021; Rezlescu et al., 2017). This particular problem with sampling error is not an issue for the subtraction approach because it is ‘only’ affected by within-subject error owing to measurement imperfection (reliability being less than perfect). It is not the case, however, that the regression approach is unaffected by within-subject error. As mentioned above, the reliabilities of two measures limits the magnitude of the correlation that can be observed between them. Thus, as compared with the subtraction approach, where each subject serves as their own control, the estimate of an individual’s performance and its residual yielded by regression is affected by both sampling error and within-subject error. In this respect, it can seem surprising that measures based on regression can turn out more reliable than measures based on subtraction (DeGutis et al., 2013; Rezlescu et al., 2017) even if this is not always the case (Ross et al., 2015). However, and as pointed out by Hedge et al. (2018), the main reason why difference scores may have low reliability is that they can be quite successful in removing between-subject variance thus increasing the amount of (within-subject) measurement error relative to between-subject variance. Given these considerations it is not clear that the regression approach necessarily represents “...an improved analytic approach” relative to an approach based on difference scores as DeGutis et al. (2013, p. 88) claim (for a deeper discussion, see Willett, 1988). In fact, DeGutis et al. did not really test whether their measure of ‘holistic’ processing was face-specific or even specific for upright faces because they did not examine whether it correlated with processing of other stimuli besides upright faces, for example objects or inverted faces (Boutet et al., 2021). The same limitation was acknowledged by Rezlescu et al. (2017) in their study. They found that a regression-based measure of the face inversion effect correlated with processing of upright faces in an independent task, but they did not test whether it also correlated with for example processing of inverted faces.

Given the issues discussed above, the present work had four objectives. We wanted to examine: (i) if a correlation can be observed between inversion effects measured with the CFMT and the CFPT when using a larger sample ($N = 420$) than the one used by Klargaard et al. (2018), (ii) whether correlations between inversion effects will be specific to tasks using faces as stimuli, or/and (iii) will depend on the approach used (subtraction vs. regression), and (iv) the degree to which inversion effects based on subtraction and regression correlate with performance with both upright and inverted faces in independent tasks. The latter objective is motivated by an interest to see how specific the inversion measures are. If they are specific to holistic processing of upright faces, the inversion measures should correlate with processing of upright but not inverted faces.

Method

Tasks The original versions of the CFMT (Duchaine & Nakayama, 2006) and the CFPT (Duchaine, Yovel, et al., 2007b) were used. To examine objective (ii)—whether potential correlations between measures of inversion effects would generalize to tasks with objects—we used the Cambridge Car Memory Test (CCMT), which is identical to the CFMT except that cars are used instead of faces as stimuli (Dennett et al., 2012). All instructions and feedback on the Cambridge tests were translated to Danish.

In the CFMT and the CCMT the participant is introduced to six target stimuli, and then tested with forced choice items consisting of three stimuli, one of which is the target. The tests comprise a total of 72 trials distributed over three phases: (a) an intro-phase with 18 trials where the study stimulus and the target stimulus are identical, (b) a novel-phase with 30 trials where the target differs from the study stimulus in pose and/or lighting, and (c) a novel+noise phase with 24 trials where the target differs from the study stimulus in pose and/or lighting and where Gaussian noise is added to the target. The dependent measure is the number of correct trials. The maximum score is thus 72; chance-level is 24 (33% correct responses).

In the CFPT, the participant must arrange six facial images according to their similarity to a target face. The images were created by morphing six different individuals with the target face. The proportion of the morph coming from the target face varies in each image (88, 76, 64, 52, 40, and 28%). The test comprises 16 trials, half with upright and half with inverted faces. Scores for each item are computed by summing the deviations from the correct position for each face. Scores for the eight trials are then added to determine the total number of respectively upright and inverted errors. Hence, the dependent measure is a deviation-score where 0 represents perfect performance and 144 the worst possible performance. Chance performance corresponds to 35% correct.

Participants A total of 420 first-year psychology students who were naïve to our hypotheses contributed data for this study as part of their course in cognitive psychology. The course is approved by the study board at the Department of Psychology, University of Southern Denmark, and the experiments conducted do not require formal ethical approval/registration according to Danish Law and the institutional requirements. Prior to participation, the students were informed that data collected in the experiments might be used in an anonymous form in future publications. Participants were free to opt-out if they wished, and participation in the experiments was taken as consent. Hence the sample size was determined by the number of students who took the course in the years 2014–2020 and provided data for all three tasks (CFMT, CCMT, and CFPT).

No participants were excluded from the analyses reported below. Task order was counterbalanced for the CFMT and the CCMT. Except for 2 years, the CFPT was always performed a week prior to the CFMT and the CCMT, which were performed on the same day. The individual data for each of the 420 participants are provided in the Supplementary Data (<https://osf.io/7ufv8/>). To comply with GDPR, age was not logged for all the participants but the majority were in the age range of 20–30 and approximately two-thirds were women.

Statistical procedures To make comparisons across tasks simple, all dependent measures were converted to percentage correct (Bate et al., 2019) and subsequent analyses were based on percentage correct responses; 95% confidence intervals were estimated based on bias corrected and accelerated bootstrapping (1000 samples) as implemented in the software package SPSS (version 28). Estimates of reliability were computed with the Spearman–Brown prediction formula (r_{sb}) also as implemented in SPSS.

Two different measures of inversion effects were computed. The first was a simple difference score obtained by subtracting performance with inverted stimuli from performance with upright stimuli. This was done for all three tasks. In the following, these measures will be referred to with the subscript ‘SUB’ for subtraction, such that $CFMT_{SUB}$ will refer for the difference score for performance with upright and inverted faces in the CFMT. The second measure was based on regressing out the effect of the control condition (inverted stimuli) from the condition of interest (upright stimuli) (DeGutis et al., 2013). The resulting standardized residual for each individual was taken as an index of ‘holistic’ processing. These measures will be referred to with the subscript ‘REG’ for regression.

As mentioned above, the correlation observed between two measures will scale with the reliabilities of the measures: The lower the reliabilities, the lower the possible correlation that can be observed between them will be. Because of this, many recommend that observed correlations be adjusted for measurement error (reliability) to give less biased estimates (Schmidt & Hunter, 1999). This can be achieved by using the disattenuation formula suggested by Spearman (1904), which is the observed correlation divided by the square root of the product of the two measures’ reliabilities. This formula will always adjust the observed correlation upward, and the lower the reliability the larger the adjustment. It is important to note that the adjusted correlation—which we will henceforth refer to as $r_{Adjusted}$ —is itself an estimate that is affected by sampling error, which will also affect the estimates of the reliabilities that it rests upon (Hedge et al., 2018). Hence, $r_{Adjusted}$ can yield values higher than 1 when reliability is low, and the adjustment in general does lead to overestimation when reliability is low (Wang, 2010). For these and related reasons, $r_{Adjusted}$ should be considered carefully (Winne & Belfry, 1982) but for

completeness we do present the adjusted correlation when a significant correlation is found (for unadjusted measures).

To address the issue of specificity—which was our last objective—we examined how well the inversion effect indexes based on the CFPT ($CFPT_{SUB}$ and $CFPT_{REG}$) correlated with processing of upright and inverted faces in the CFMT, and how well the inversion effect indexes based on the CFMT ($CFMT_{SUB}$ and $CFMT_{REG}$) correlated with processing of upright and inverted faces in the CFPT. While we could have addressed specificity with another set of face processing tasks that were not also used to derive the inversion effect indexes from, we believe these comparisons are nevertheless fair because: (i) they are based on independent tasks (the indexes were derived from another task than the test conditions), and (ii) holistic processing is thought of as a perceptual phenomenon (Rezlescu et al., 2017) and both the CFMT and the CFPT require perceptual processing. In the following we will use the term “benchmark” to refer to these tests of specificity. Hence, the CFMT will serve as the benchmark task for evaluating the specificity of the $CFPT_{SUB}$ and $CFPT_{REG}$ measures and the CFPT will serve as benchmark task for evaluating the specificity of the $CFMT_{SUB}$ and $CFMT_{REG}$ measures.

Results

Inversion effects As can be seen in Table 1, the mean performance with upright stimuli was highest for the CFMT (81% correct), intermediate for the CFPT (77% correct) and lowest for the CCMT (68% correct). These figures are similar to previous reports with similar age groups (see Table 2). Compared with the performance with upright stimuli, performance with inverted stimuli was rather similar across the three tasks ranging from 58–62% correct. Table 1 also shows there were credible inversion effects across all tasks (the 95% CI’s for reduction in % correct scores as a consequence of inversion did not contain 0), and the magnitude of these effects—around 20% for faces and 6% for objects—are comparable with what has been reported for faces and objects in previous studies (McKone & Robbins, 2011; Rezlescu et al., 2017). Hence, the inversion effects were disproportionately larger for faces than for objects which is the typical finding (Bruyer, 2011). Finally, because correlations can be affected by range restrictions, it is also worth noting that the coefficients of variation are rather similar across the six conditions (range 13–18) (see Table 2) and again comparable with previous findings (see e.g., Gauthier, 2018). Information regarding all correlations between tasks and conditions with upright and inverted stimuli can be found in the Appendix Table 6.

Reliabilities The split-half reliabilities of the measurements were generally moderate to good for the CFMT and the CCMT (r_{sb} ranged from .57 to .88) (see Table 3). However,

Table 1 Mean performance (% correct) with upright and inverted stimuli in the three tests and mean reduction in performance in the tests as a function of inversion. The 95% CIs of the means are given in brackets

	Upright % correct Mean / 95% CI	Inverted % correct Mean / 95%CI	Reduction % correct (Upright - Inverted) Mean / 95% CI
Cambridge Face Memory Test	80.6 [79.6, 81.6]	59.4 [58.6, 60.2]	21.2 [20.2, 22.2]
Cambridge Face Perception Test	76.6 [75.6, 77.6]	57.9 [57, 58.9]	18.7 [17.8, 19.7]
Cambridge Car Memory Test	67.9 [66.8, 68.9]	61.6 [60.5, 62.6]	6.3 [5.2, 7.2]

Table 2 The coefficient of variation for the six conditions in the present study and from other studies that have used the same test conditions in populations with a similar age range

	CFMT Upright	CFMT Inverted	CFPT Upright	CFPT Inverted	CCMT Upright	CCMT Inverted	<i>N</i>	Age range
Present study	14 (M = 81, SD = 11)	15 (M = 59, SD = 9)	13 (M = 77, SD = 10)	17 (M = 58, SD = 10)	16 (M = 68, SD = 11)	18 (M = 62, SD = 11)		
Bate et al. (2019)	15 (M = 81, SD = 12)	-	12 (M = 77, SD = 9)	-	-	-	62	18–34
Bowles et al. (2009)	16 (M = 77, SD = 12)	-	9 (M = 75, SD = 7)	15 (M = 57, SD = 9)	-	-	124 CFMT / 36 CFPT	18–35
Dennett et al. (2012)	-	-	-	-	22 (M = 74, SD = 16)	-	135	18–32
Gray et al. (2019)	14 (M = 84, SD = 12)	-	-	-	23 (M = 74, SD = 17)	-	61	M = 37 / SD = 10
Klargaard et al. (2018)	12 (M = 84, SD = 10)	17 (M = 59, SD = 10)	12 (M = 73, SD = 9)	15 (M = 57, SD = 9)	14 (M = 73, SD = 10)	16 (M = 68, SD = 11)	32	M = 32 / SD = 10
Rezlescu et al. (2017)	-	-	14 (M = 74, SD = 10)	19 (M = 54, SD = 10)	-	-	80	M = 20.5 / SD = 2.1
Wilmer et al. (2010)	17 (M = 77, SD = 13)	-	-	-	-	-	557	20–35

In brackets are given the mean % correct (M) and the standard deviation (SD) for the condition. *N* is the number of participants in the given study (CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test; CCMT = Cambridge Car Memory Test)

Table 3 The Spearman–Brown split-half reliabilities associated with the measures used in the present study

	Upright condition	Inverted condition	Difference score Upright - Inverted	Residuals Upright vs. Inverted
Cambridge Face Memory Test	.88	.69	.66	.81
Cambridge Face Perception Test	.54	.44	-.05	.32
Cambridge Car Memory Test	.80	.76	.57	.67

the reliability of measurements based on the CFPT were considerably lower and for the CFPT_{SUB} reliability could not be computed because of a negative correlation between odd and even trials. Finally, for all measures of inversion effects, the estimates based on regression were more reliable than the ones based on difference scores.

Correlations between tasks in inversion effects There was no significant correlation between the CFPT_{SUB} and the CFMT_{SUB} measures of inversion effects (see Table 4). However, a small significant correlation was found between the CFMT_{SUB} and the CCMT_{SUB} measures ($r = .15$, $r_{\text{Adjusted}} = .24$). For the measures based on regression significant

Table 4 The correlations between tasks in inversion effects based on either subtraction (SUB) or regression (REG) measures

	CFMT _{SUB}	CFPT _{SUB}	CCMT _{SUB}	CFMT _{REG}	CFPT _{REG}
CFPT _{SUB}	.04 [-.05, .14]				
CCMT _{SUB}	.15 [.06, .24] (.24)	.03 [-.08, .15]			
CFMT _{REG}	-	-	-		
CFPT _{REG}	-	-	-	.12 [.02, .21] (.23)	
CCMT _{REG}	-	-	-	.11 [.01, .21] (.14)	.01 [-.09, .11]

The 95% CIs of the correlations are given in brackets. Significant effects are marked in bold. For significant correlations, the disattenuated correlations are given in parentheses. (CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test; CCMT = Cambridge Car Memory Test)

Table 5 Correlations between the different measures of inversion effects (SUB = subtraction score; REG = regression residual) and performance with upright and inverted stimuli in the Cambridge Face Memory Test (CFMT), the Cambridge Face Perception test (CFPT), and the Cambridge Car Memory Test (CCMT)

	CFMT _{SUB}	CFPT _{SUB}	CFMT _{REG}	CFPT _{REG}
CFMT Upright	-	-.02 [-.11, .08]	-	.17 [.07, .27] (.31)
CFMT Inverted	-	-.07 [-.17, .02]	-	.13 [.04, .23] (.29)
CFPT Upright	.06 [-.03, .16]	-	.18 [.08, .28] (.27)	-
CFPT Inverted	.02 [-.09, .12]	-	.16 [.05, .25] (.26)	-

95% CI of the correlations are given in brackets. Significant correlations are marked in bold. For significant correlations, the disattenuated correlations are given in parentheses

correlations were found both between the CFMT_{REG} and the CFPT_{REG} measures ($r = .12, r_{Adjusted} = .23$) and between the CFMT_{REG} and the CCMT_{REG} measures ($r = .11, r_{Adjusted} = .14$). The magnitudes of the two correlations found with the regression measures did not differ significantly ($Z = .15, p = .88$). This also held true when the comparison was performed on the adjusted correlations ($Z = 1.3, p = .18$). Likewise, the correlation between the CFMT_{SUB} and the CCMT_{SUB} measures was not significantly higher than the correlation found between the CFMT_{REG} and the CCMT_{REG} measures ($Z = .59, p = .56$: for the adjusted correlation $Z = 1.5, p = .13$).

Specificity None of the measures based on difference scores (CFPT_{SUB} and CFMT_{SUB}) correlated significantly with processing of upright or inverted faces in the benchmark tasks (see Table 5). In comparison, the CFMT_{REG} measure correlated with processing of upright faces in the CFPT ($r = .18, r_{Adjusted} = .27$) and the CFPT_{REG} measure correlated with processing of upright faces in the CFMT ($r = .17, r_{Adjusted} = .31$). Hence, both regression measures of inversion effects were related to processing of upright faces in the benchmark tasks. However, the same measures were also related to processing of inverted faces (CFMT_{REG} and CFPT_{Inverted} $r = .16, r_{Adjusted} = .26$; CFPT_{REG} and CFMT_{Inverted} $r = .13, r_{Adjusted} = .29$) and the magnitude of these correlations did not differ considerably from correlations found with upright faces (cf. the 95% CIs in Table 5). In other words, the more poorly performance in the control condition (e.g., CFMT_{Inverted}) is at predicting performance in the condition of interest (e.g., CFMT_{Upright}),

the more likely it is that the individual will obtain a high score on the benchmark task (e.g., the CFPT) regardless of whether the stimuli are upright or inverted.

To assess how systematic the correlations between the regression measures and performance on the benchmark tasks were, we examined whether the individuals who drove the correlation between CFPT_{REG} and CFMT_{Upright} performance were the same individuals who drove the correlation between CFPT_{REG} and CFMT_{Inverted} performance. To do so, we computed the residuals from the correlation between the CFPT_{REG} and CFMT_{Upright} performance on the one hand (CFPT_{REG}+CFMT_{Upright Residual}), and the residuals from the correlation between the CFPT_{REG} and CFMT_{Inverted} performance on the other (CFPT_{REG}+CFMT_{Inverted Residual}). We then performed a correlation analysis with these two variables (CFPTREG+CFMTUpright Residual & CFPTREG+CFMTInverted Residual). This yielded a moderate to large effect ($r = .47, 95\% CI [.39, .54]$). A similar sized effect was found when we looked at the relationship between the CFMT_{REG} and CFPT performance with upright and inverted faces respectively (CFMT_{REG}+CFPT_{Upright Residual} & CFMT_{REG}+CFPT_{Inverted Residual}: $r = .49, 95\% CI [.4, .56]$). These results reveal some consistency across tasks. It is to a considerable degree the same individuals who drive the (modest) correlations observed between the CFPT_{REG} and performance with of upright and inverted faces on the CFMT, and also the same individuals who drive the (modest) correlations observed between the CFMT_{REG} and performance with upright and inverted faces on the CFPT.

Discussion

Our first objective was to examine whether inversion effects derived from different tasks measure the same construct. We found some evidence in favor of this proposition in that inversion effects derived from the Cambridge Face *Memory* Test (CFMT) did correlate with inversion effects derived from the Cambridge Face *Perception* Test (CFPT). This, however, was true only when the inversion effects were estimated by means of regression (CFMT_{REG} & CFPT_{REG}), and not when estimated by means of subtraction (CFMT_{SUB} and CFPT_{SUB}). The correlation based on regression was rather small, however, and the two measures shared only 1.4% of their variances. Even in the hypothetical event that the measures had had perfect reliability this would only have amounted to 5% of shared variance; an adjusted measure based on dividing the correlation observed by the product of the reliabilities of its two components (Spearman, 1904). Finally, the CFMT_{REG} measure shared a comparable amount of variance (1.1%, 2%_{Adjusted}) with the regression measure of inversion effects based on the Cambridge *Car* Memory Test (CCMT_{REG}). A similar picture was obtained for measures based on subtraction where the CFMT_{SUB} measure shared 2.1% (5.7%_{Adjusted}) of its variance with the CCMT_{SUB} measure. Given these findings there is limited evidence to support the notion that inversion effects derived from different tasks measure the same construct to any great extent or that they are category-specific. Also, the consistency across the two methods (subtraction and regression) was mixed in that the correlation between inversion effects for the CFMT and the CCMT was found with measures based on both regression and subtraction, whereas the correlation between inversion effects for the CFMT and the CFPT was found only with measures based on regression.

The finding that measures of inversion effects shared a comparable, but small, amount of variance across tasks with faces and objects is troubling if it is assumed that faces, and faces alone, are subject to holistic processing. Many researchers, however, take the stance that holistic processing is not limited to faces (Gerlach et al., 2022) but may be of greater importance for face than for object processing (e.g., McKone & Robbins, 2011). The present findings are compatible with such an interpretation as the drop in performance due to inversion was more pronounced for faces (around 20%) than for objects (6%). It is somewhat more troubling that the inversion effects observed are not more tightly coupled for faces in the two face processing tasks than they are for faces and objects in the recognition tasks. After all, the reductions in performance due to inversion were of similar magnitude for the CFMT (21%) and the CFPT (19%). It is also troubling, as we will discuss below, that the inversion indexes for faces correlate equally well

with upright and inverted faces when based on the regression method.

As argued in the Introduction, it is generally assumed that the superior performance with upright compared with inverted faces reflects that upright (but not inverted) faces are processed holistically. Consequently, to the degree that measures of inversion effects do indeed provide indexes of holistic processing we would expect these measures to yield higher correlations with processing of upright faces—that according to theory are subject to holistic processing—than with processing of inverted faces that—again according to theory—are *not* subject to holistic processing but are processed in a parts-based manner (Piepers & Robbins, 2012; Rossion, 2008).

None of the measures based on subtraction revealed any significant correlation with processing of upright or inverted faces in the benchmark tasks. In comparison, both the CFPT_{REG} and the CFMT_{REG} accounted for a small but significant portion of the variance in processing of upright faces in the benchmark tasks (3%: 10%_{Adjusted} and 7%_{Adjusted} respectively for the CFPT_{REG} and the CFMT_{REG}). In this respect, the present findings add to an increasing number of studies which find that measures of holistic processing can only explain a modest degree of the variation in face processing ability observed among individuals if any at all (Konar et al., 2010; Rezlescu et al., 2017; Richler et al., 2015; Richler & Gauthier, 2014; Verhallen et al., 2017). Moreover, and perhaps more critically, the measures accounted for a similar amount of variance with inverted faces in the benchmark tasks (2%: 8%_{Adjusted} and 7%_{Adjusted} respectively, for the CFPT_{REG} and the CFMT_{REG}) as they did with upright faces. Considered together, these findings suggests that while both the CFPT_{REG} and CFMT_{REG} measures can account for some of the variance found in other tasks of face processing, there is no evidence suggesting that these measures isolate something which is specific to upright faces. It is worth noting that even if these results may have been affected by the low reliability of the CFPT, this alone cannot account for the present findings. First of all, the regression-based measure derived from the CFPT did reveal credible correlations with the CFMT despite the measure's poor reliability. Secondly, adjusting the correlations for the reliability of the measures did not change the pattern observed: The amount of variance explained in the benchmark tasks was quite similar for upright and inverted faces. This orientation-invariance also applies to overall task performance. The amounts of shared variance between the CFMT_{Upright} and the CFPT_{Upright} on the one hand and between the CFMT_{Upright} and the CFPT_{Inverted} on the other were quite similar ($r = .3$ and $r = .31$, respectively; see the Appendix Table 6). The same was the case for the CFPT_{Upright} and the CFMT_{Upright}/CFMT_{Inverted} ($r = .3$ and $r = .3$ respectively, see the Appendix Table 6). For similar findings regarding orientation-invariance, see Meinhardt et al. (2019).

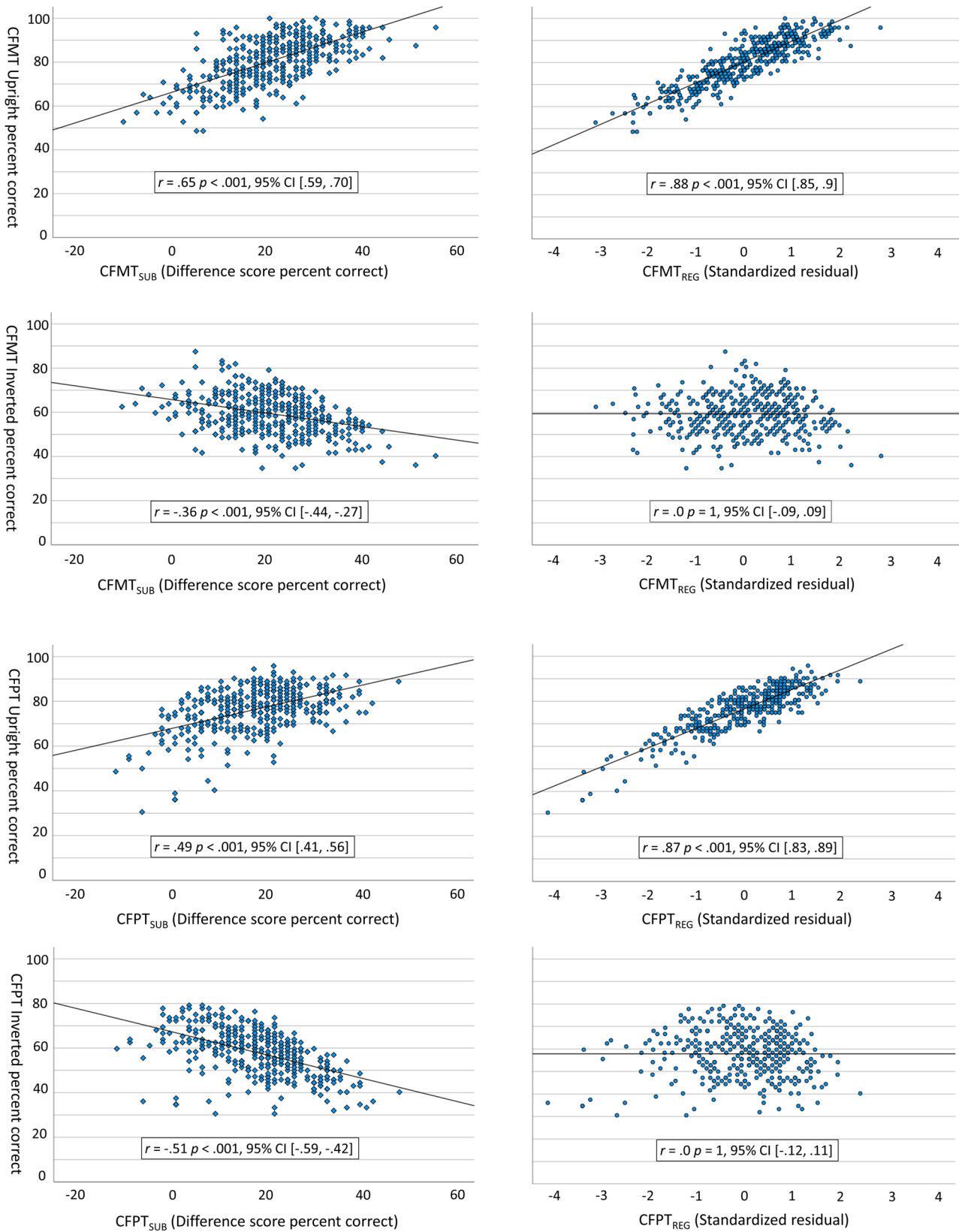


Fig. 1 Scatter plots showing how the two types of indexes based on subtraction (*left panels*) and regression (*right panels*) correlated with their constituent conditions. As can be seen, the indexes based on subtraction correlated with both the control condition (inverted stim-

uli) and the condition of interest (upright stimuli). In comparison, the indexes based on regression yielded more pure measures reflecting variation in performance with upright but not inverted stimuli

The finding that regression measures are apparently *insensitive* to the orientation of the stimuli in the benchmark tasks seems at odds with the claim by DeGutis et al. (2013) that regression yield specific measures of (holistic) processing. Admittedly, this finding of lack of specificity is surprising, as there can be no doubt that regression does ensure that the control condition (inverted faces) does not correlate with the measure of inversion effects ($CFMT_{REG}/CFPT_{REG}$)—which presumably reflects holistic processing—whereas the condition of interest (upright faces) does. In comparison, difference scores ($CFMT_{SUB}/CFPT_{SUB}$) may be affected by both the control condition and the condition of interest. As can be seen in Fig. 1, this is also the case with our data. Hence, as also demonstrated by DeGutis et al. (2013), the regression-based method does yield a somewhat more specific measure than the method based on subtraction in that the regression measure based on for example the CFMT only correlates with processing of upright stimuli and not inverted stimuli in the CFMT whereas the subtraction measure based on the CFMT correlates with processing of *both* upright and inverted stimuli in the CFMT. However, and as shown here, the regression measures are only specific in the context of the tasks that they were derived from, that is, *the specificity does not generalize to other tasks*. Consequently, even though the regression-based measure derived from the CFMT only correlated with processing of upright and not inverted stimuli in the CFMT it correlated with processing of both upright and inverted stimuli in the CFPT, and to similar degrees. These findings raise serious concerns as to what exactly is captured by these regression measures. Even though we cannot answer this question because the individual difference approach adopted here rests on common variance, what we can conclude is that it is not holistic processing if holistic processing is defined as processing that is limited to upright faces.

If the $CFMT_{REG}$ and $CFPT_{REG}$ do not measure holistic processing, does this mean that faces are not processed holistically? Probably not. It is entirely likely, as some have proposed, that both upright and inverted faces are processed holistically (Meinhardt et al., 2019; Murphy & Cook, 2017; Murphy et al., 2020; Sekuler et al., 2004). If this is indeed the case, then variance due to holistic processing will of course not be captured by the residuals in the $CFMT_{REG}$ and $CFPT_{REG}$ measures; only operations that differ between processing of upright and inverted faces will. This would certainly help explain why the $CFMT_{REG}$ and the $CFPT_{REG}$ measures do not seem sensitive to orientation across tasks. It is worth noting, though, that this suggestion does not mean that upright and inverted faces are processed equally well. Clearly, there are huge differences also in the present sample between the participants' performance with upright and inverted faces (a 20% reduction due to inversion). It simply means that individual differences in holistic processing will

contribute equally to performance differences across subjects in processing of upright and inverted faces.

Regardless of whether the $CFMT_{REG}$ and $CFPT_{REG}$ measures gauge holistic processing or not—and we are inclined to say ‘not’—what they measure seems rather task-dependent, as they could only account for about 2–3% (7–10%_{Adjusted}) of the variance in performance with upright and inverted faces in the benchmark tasks. A similar figure was observed for the amount of variance shared by the measures directly ($CFPT_{REG}$ and $CFMT_{REG}$), which was 1.4% (5%_{Adjusted}). To this we must add the observation that the $CFMT_{REG}$ also shared variance with the $CCMT_{REG}$ (1%, 2%_{Adjusted}). The only conclusion we can reach based on these findings is that inversion effects are quite task-specific and do not generalize better among face processing tasks than they do across face and object (car) processing tasks. Further, given that a similar amount of shared variance was found between the $CFMT_{SUB}$ and $CCMT_{SUB}$ measures (2%, 6%_{Adjusted}) this finding is stable across both regression and subtraction.

Conclusions

We find clear inversion effects for faces in two widely used tests in the neuropsychological literature—the Cambridge Face Perception Test and the Cambridge Face Memory Test—and the magnitude of reduction due to inversion in these tests is much larger (20%) than what is observed for objects (cars) in the Cambridge Car Memory Test (6% reduction). However, the inversion effects were quite task-dependent, and did not correlate better among face processing tasks than across face and object processing tasks. Finally, in contrast to previous studies, we also explicitly tested whether the measures of inversion effects provided specific measures of holistic processing—that is, tapped into operations confined to processing of upright faces. We found no evidence supporting this proposition. In conclusion, the present findings suggest that inversion effects are highly context-dependent and do not reflect holistic processing if holistic processing is defined as processing that is specific to upright faces. These observations are done in a much larger sample than previous studies ($N = 420$) and are thus unlikely to be caused by low power or sampling error.

The present findings do not bring us closer to why faces are often much more affected by inversion than objects are or whether inversion leads to qualitative or quantitative shifts in processing. They do suggest, however, that we cannot treat the inversion effect with faces as a pure or generalizable measure of holistic face processing. This is an important lesson, not least for studies that wish to understand what inversion effects reflect.

Appendix

Table 6 Correlations between all conditions. CFMT = Cambridge Face Memory Test, CFPT= Cambridge Face Perception Test, and CCMT = Cambridge Car Memory Test. 95% CI of the correlations are given in brackets

	CFMT Upright	CFMT Inverted	CFPT Upright	CFPT Inverted	CCMT Upright
CFMT Inverted	.48 [.41, .56]				
CFPT Upright	.30 [.21, .39]	.30 [.21, .38]			
CFPT Inverted	.31 [.22, .40]	.37 [.28, .45]	.50 [.41, .58]		
CCMT Upright	.14 [.05, .25]	.14 [.05, .23]	.11 [.01, .21]	.15 [.06, .26]	
CCMT Inverted	.15 [.05, .25]	.31 [.22, .39]	.16 [.05, .26]	.23 [.14, .32]	.55 [.48, .62]

Code availability Not applicable

Funding This work was supported by the Independent Research Fund Denmark [DFP – 4001-00115] to Christian Gerlach. We wish to express our gratitude to the Friends of Fakutsi Association (FFA).

Declarations

Conflicts of interest/Competing interests The authors declare no conflict of interest

Ethics approval The experiments conducted do not require formal ethical approval/registration according to Danish law.

Consent to participate Participants were informed orally and in writing that participation was a voluntary part of their course in cognitive psychology and that they were free not to participate if they so wished. Participation in the experiments was taken as consent.

Consent for publication Participants were informed orally and in writing that data collected in the experiments might be used in an anonymous form in future publications.

References

- Bate, S., Bennetts, R. J., Gregory, N., Tree, J. J., Murray, E., Adams, A., Bobak, A. K., Penton, T., Yang, T., & Banissy, M. J. (2019). Objective patterns of face recognition deficits in 165 adults with self-reported developmental prosopagnosia. *Brain Sciences*, 9(6). <https://doi.org/10.3390/brainsci9060133>
- Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: A case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology*, 34(6), 357–376.
- Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, 30(1), 81–91.
- Boutet, I., Nelson, E. A., Watier, N., Cousineau, D., Béland, S., & Collin, C. A. (2021). Different measures of holistic face processing tap into distinct but partially overlapping mechanisms. *Attention, Perception, & Psychophysics*, 83(7), 2905–2923. <https://doi.org/10.3758/s13414-021-02337-7>
- Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C. E., & 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), 423–455. <https://doi.org/10.1080/02643290903343149>
- Bruyer, R. (2011). Configural face processing: A meta-analytic survey. *Perception*, 40(12), 1478–1490.
- Cashon, C. H., & Holt, N. A. (2015). Developmental origins of the face inversion effect. *Advances in Child Development and Behavior*, 48, 117–150. <https://doi.org/10.1016/bs.acdb.2014.11.008>
- DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, 126(1), 87–100. <https://doi.org/10.1016/j.cognition.2012.09.004>
- 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. <https://doi.org/10.3758/s13428-011-0160-2>
- Duchaine, B., Germine, L., & Nakayama, K. (2007a). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419–430.
- 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(4), 576–585. <https://doi.org/10.1016/j.neuropsychologia.2005.07.001>
- Duchaine, B., Yovel, G., & Nakayama, K. (2007b). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social Cognitive and Affective Neuroscience*, 2(2), 104–113. <https://doi.org/10.1093/scan/nsm003>
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105(3), 482–498.
- Gauthier, I. (2018). Domain-specific and domain-general individual differences in visual object recognition. *Current Directions in Psychological Science*, 27(2), 97–102. <https://doi.org/10.1177/0963721417737151>
- Gerlach, C., Kühn, C. D., Mathiassen, A. B., Kristensen, C. L., & Starfelt, R. (2022). The face inversion effect or the face upright

- effect? *Cognition*, 232, 105335. <https://doi.org/10.1016/j.cognition.2022.105335>
- Gray, K. L. H., Biotti, F., & Cook, R. (2019). Evaluating object recognition ability in developmental prosopagnosia using the Cambridge car memory test. *Cognitive Neuropsychology*, 1–8. <https://doi.org/10.1080/02643294.2019.1604503>
- Griffin, J. W. (2020). Quantifying the face inversion effect in nonhuman primates: A phylogenetic meta-analysis. *Animal Cognition*, 23(2), 237–249. <https://doi.org/10.1007/s10071-019-01340-8>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Hendel, R. K., Starrfelt, R., & Gerlach, C. (2019). The good, the bad, and the average: Characterizing the relationship between face and object processing across the face recognition spectrum. *Neuropsychologia*, 124, 274–284. <https://doi.org/10.1016/j.neuropsychologia.2018.11.016>
- James, T. W., Arcurio, L. R., & Gold, J. M. (2013). Inversion effects in face-selective cortex with combinations of face parts. *Journal of Cognitive Neuroscience*, 25(3), 455–464. https://doi.org/10.1162/jocn_a_00312
- Klargaard, S. K., Starrfelt, R., & Gerlach, C. (2018). Inversion effects for faces and objects in developmental prosopagnosia: A case series analysis. *Neuropsychologia*, 113, 52–60. <https://doi.org/10.1016/j.neuropsychologia.2018.03.026>
- Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, 21(1), 38–43. <https://doi.org/10.1177/0956797609356508>
- Leder, H., Goller, J., Forster, M., Schlageter, L., & Paul, M. A. (2017). Face inversion increases Attractiveness. *Acta Psychol (Amst)*, 178, 25–31. <https://doi.org/10.1016/j.actpsy.2017.05.005>
- McKone, E., & Robbins, R. (2011). Are faces special? In G. Rhodes, A. Calder, M. Johnson, & J. V. Haxby (Eds.), *Oxford handbook of face perception* (pp. 149–176). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199559053.013.0009>
- Meinhardt, G., Meinhardt-Injac, B., & Persike, M. (2019). Orientation-invariance of individual differences in three face processing tasks. *Royal Society Open Science*, 6(1), –181350. <https://doi.org/10.1098/rsos.181350>
- Murphy, J., & Cook, R. (2017). Revealing the mechanisms of human face perception using dynamic apertures. *Cognition*, 169, 25–35. <https://doi.org/10.1016/j.cognition.2017.08.001>
- Murphy, J., Gray, K. L. H., & Cook, R. (2020). Inverted faces benefit from whole-face processing. *Cognition*, 194, 104105. <https://doi.org/10.1016/j.cognition.2019.104105>
- Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, 19(4), 655–662.
- Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, 3, 559. <https://doi.org/10.3389/fpsyg.2012.00559>
- Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology. Human Perception and Performance*, 43(12), 1961–1973. <https://doi.org/10.1037/xhp0000400>
- Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, 47(1), 25–57. [https://doi.org/10.1016/0010-0277\(93\)90061-y](https://doi.org/10.1016/0010-0277(93)90061-y)
- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, 22(4), 464–471. <https://doi.org/10.1177/0956797611401753>
- Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, 15(9), 15. <https://doi.org/10.1167/15.9.15>
- Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, 140(5), 1281–1302. <https://doi.org/10.1037/a0037004>
- Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology*, 3, 553. <https://doi.org/10.3389/fpsyg.2012.00553>
- Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite-task measurements of holistic face processing. *Behavior Research Methods*, 47(3), 736–743. <https://doi.org/10.3758/s13428-014-0497-4>
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, 128(2), 274–289. <https://doi.org/10.1016/j.actpsy.2008.02.003>
- Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, 21(2), 139–253.
- Rossion, B. (2015). Face perception. *Brain Mapping: An Encyclopedic Reference*, 2, 515–522.
- Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, 8(4), 3.1–13. <https://doi.org/10.1167/8.4.3>
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252–257. <https://doi.org/10.3758/pbr.16.2.252>
- Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence*, 27(3), 183–198. [https://doi.org/10.1016/S0160-2896\(99\)00024-0](https://doi.org/10.1016/S0160-2896(99)00024-0)
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5), 391–396. <https://doi.org/10.1016/j.cub.2004.02.028>
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal Experimental Psychology A*, 46(2), 225–245.
- Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic face processing. In A. J. Calder, G. Rhodes, M. K. Johnson, & J. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 177–194). Oxford University Press.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25(5), 583–592. <https://doi.org/10.3758/bf03211301>
- Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *Quarterly Journal Experimental Psychology (Hove)*, 69(10), 1876–1889. <https://doi.org/10.1080/17470218.2016.1146780>
- Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, 141, 217–227. <https://doi.org/10.1016/j.visres.2016.12.014>
- Wang, L. L. (2010). Disattenuation of correlations due to fallible measurement. *Newborn and Infant Nursing Reviews*, 10(1), 60–65. <https://doi.org/10.1053/j.nainr.2009.12.013>

- Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. <https://doi.org/10.1177/0956797611420575>
- Willett, J. B. (1988). Chapter 9: Questions and answers in the measurement of change. *Review of Research in Education*, *15*(1), 345–422. <https://doi.org/10.3102/0091732x015001345>
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. <https://doi.org/10.1073/pnas.0913053107>
- Winne, P. H., & Belfry, M. J. (1982). Interpretive problems when correcting for attenuation. *Journal of Educational Measurement*, *19*(2), 125–134 <http://www.jstor.org/stable/1434905>
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. <https://doi.org/10.1037/h0027474>
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759. <https://doi.org/10.1068/p160747>

Open Practices Statement The data supporting the findings of this study are available within the article and its supplementary materials. None of the experiments were preregistered.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.