Beliefs alter holistic face processing … if response bias is not taken into account

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

Olivia S. Cheung Martinos Center, Massachusetts General Hospital, Harvard Medical School, USA

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

The composite paradigm is widely used to quantify holistic processing (HP) of faces, but there is debate regarding the appropriate design (partial vs. complete) and measures in this task. Here, we argue that some operational definitions of HP are problematic because they are sensitive to top-down influences, even though the underlying concept is assumed to be cognitively impenetrable. In Experiment 1, we told one group of participants that the target face half would remain the same on 75% of trials and another group that it would change on 75% of trials. The true proportion of same/different trials was 50%—groups only differed in their beliefs about the target halves. In Experiment 2, we manipulated the actual proportion of same/different trials in the experiment (75% of trials were the same for one group; 75% of trials were different for another group) but did not give explicit instructions about proportions. In both experiments, these manipulations influenced response biases that altered partial design measures of HP while the complete design measure was unaffected. We argue that the partial design should be abandoned because it has poor construct validity.

Keywords: face recognition, holistic processing

Citation: Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Beliefs alter holistic face processing … if response bias is not taken into account. *Journal of Vision, 11*(13):17, 1–13, http://www.journalofvision.org/content/11/13/17, doi:10.1167/11.13.17.

Introduction

Holistic processing (HP) is considered a hallmark of face perception (McKone, Kanwisher, & Duchaine, 2007): Faces are not decomposed into parts or features but rather are processed as single wholes. However, HP is an unobservable construct created in attempts to account for differences between experimental conditions in behavioral experiments. In a landmark paper, Young, Hellawell, and Hay (1987) inferred that faces are processed holistically based on the finding that naming latencies for a face half were influenced by whether it was aligned or misaligned with a face half belonging to a different identity. Since then, a sequential matching version of this composite task has gained popularity as a means to measure HP. Participants match one-half of sequentially presented composite faces, made out of the top and bottom halves of different faces, and evidence that task-irrelevant parts influence performance is used as an indicator of HP: Participants cannot selectively attend because faces are processed as wholes. When the meaningful face configuration is disrupted (e.g., by misaligning the face halves), this interference is reduced (e.g., Le Grand, Mondloch, Maurer, & Brent, 2004; Richler, Tanaka, Brown, & Gauthier, 2008).

However, there are currently two different versions of the composite task in the literature and there is debate over which is more appropriate (e.g., Gauthier & Bukach, 2007; McKone & Robbins, 2007). This debate helps refine our understanding of the construct of HP at the same time as we are trying to perfect its measurement, a normal process in the study of psychological constructs, whereby a measure is evaluated through its ability to relate to and mediate between other theoretical and empirical constructs (Cronbach & Meehl, 1955). In the present case, important theoretical conclusions about normal face processing, its development, face processing deficits, and the expertise hypothesis hinge on which measure of HP is ultimately adopted as a standard because the two measurement approaches often yield qualitatively different results (e.g., Cheung, Richler, Palmeri, & Gauthier, 2008; Robbins & McKone, 2007 vs. Wong, Palmeri, & Gauthier, 2009). For instance, using one measure of HP leads to the conclusion that HP is not correlated with face recognition (*partial design*, Konar, Bennett, & Sekuler, 2010), whereas using the other measure of HP reveals such a correlation (*complete design*, Richler, Cheung, & Gauthier, 2011). Since Richler et al. replicated Konar et al.'s (2010) findings when they analyzed their results using only the partial design trials of their complete design experiment, this appears to reflect a difference in

doi: 10.1167/11.13.17 Received May 25, 2011; published November 18, 2011 ISSN 1534-7362 © ARVO

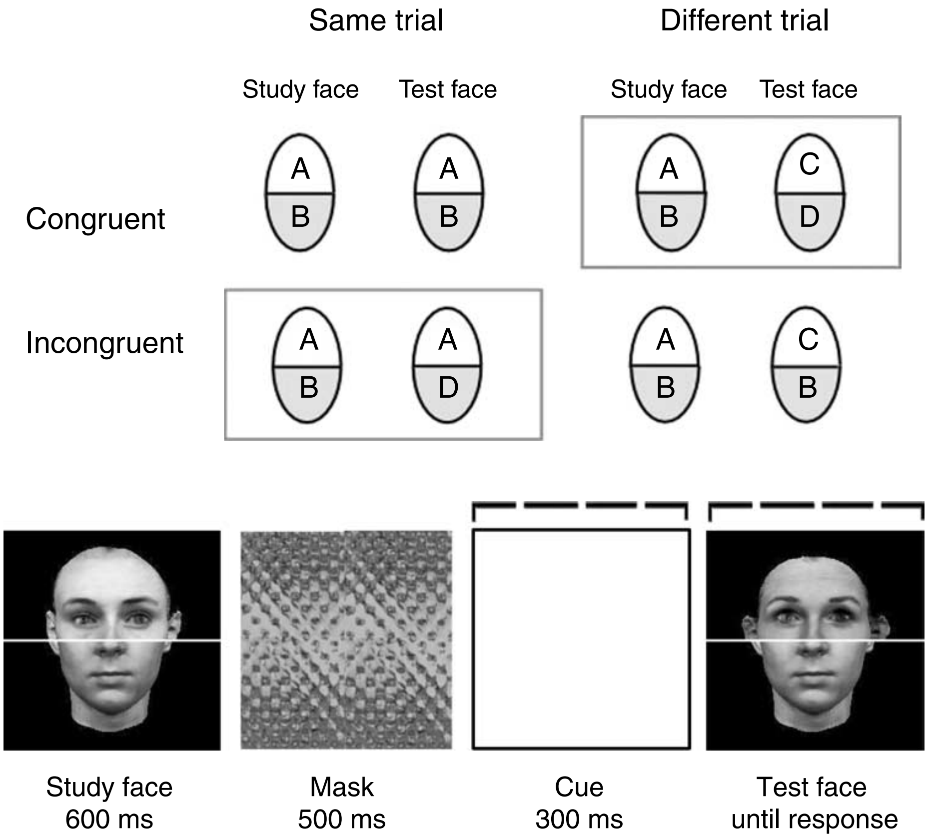

Figure 1. Design of the composite task and sample trial structure. In the schematic diagram (top panel), letters represent facial identities. Target face halves are shown in white, and task-irrelevant halves are shown in gray. In *same* trials, target halves of the study and test faces were the same; in *different* trials, target halves were different. Both types of trials featured congruent and incongruent conditions in the complete design. In congruent *same* trials, the irrelevant halves of the study and test faces were the same; in incongruent *same* trials, the irrelevant halves were different. In congruent *different* trials, the irrelevant halves of the study and test faces were different; in incongruent *different* trials, the irrelevant halves were the same. Study face halves were presented aligned, and test face halves could be either aligned or misaligned. In the partial design version of this task, only the trial types outlined in the gray boxes are presented; in the complete design, all trial types are presented. The example (bottom panel) illustrates the stimuli and trial sequence used in this study.

measures and not in other aspects of the studies (see also Cheung et al., 2008).

Here, we seek to determine which—if any—of these measures is appropriate for indexing HP by considering the influence of top-down biases. The two competing operational definitions of HP are described below.

In both versions of the composite task, participants are asked to judge whether one-half (e.g., top) of a test face is the same as or different from the corresponding half of a study face. For the remainder of the manuscript, "same" and "different" refer to the relationship between the study and test halves. Schematics of the trial types used in both designs of the composite task are illustrated in Figure 1. In the *partial design* of the composite task (trials in gray boxes in Figure 1, top panel), the irrelevant face half is always "different" and an *alignment effect* indexes HP: accuracy on "same" trials (where the target face half is the "same") is often greater for misaligned than aligned trials (e.g., Goffaux & Rossion, 2006; Le Grand et al., 2004; Rossion & Boremanse, 2008). When the meaningful face configuration is disrupted by misaligning face halves, HP is reduced, resulting in less interference from the irrelevant face half.

In the *complete design*, all partial design trials are included, plus trials where the irrelevant halves are the same. This allows definition of a factor called congruency. The irrelevant face half can be associated with the same correct response as the target half (both test face halves are the same as or different from the study face; these are congruent trials) or the irrelevant face half can be associated with a different correct response than the target half (e.g., one face half is the same as the study face, and the other face half is different from the study face; these are incongruent trials). In this design, HP is indexed by a *congruency effect*: Performance is better on congruent than on incongruent trials (Cheung et al., 2008; Farah, Wilson, Drain, & Tanaka, 1998), indicating a failure of selective attention. This congruency effect is reduced with misalignment (Cheung et al., 2008; Richler et al., 2008), and the interaction between congruency and alignment is considered more diagnostic of interference that arises due to expertise-driven HP (Richler, Wong, & Gauthier, 2011).

One issue with the partial design is that HP is often measured in terms of hit rate (percentage of correct "same" responses on trials where the irrelevant parts are "different"), so differences in response bias for aligned vs. misaligned trials (Cheung et al., 2008; Richler, Mack, Palmeri, & Gauthier, 2011; Richler et al., 2008) could influence performance. In contrast, performance in the complete design can be measured in terms of sensitivity (d'), which is independent of response bias. However, it is not just response biases related to alignment that are a concern in the partial design. As can be appreciated from Figure 1, in the partial design, irrelevant parts are always "different," so "same" trials are always incongruent and "different" trials are always congruent. This is an issue because several studies have shown that congruency itself influences response bias (Cheung et al., 2008; Richler, Mack, Gauthier, & Palmeri, 2009; Richler, Mack et al., 2011; Richler et al., 2008).[1] Moreover, spatial frequency filtering (Cheung et al., 2008) and orientation (Richler, Mack et al., 2011) have also been shown to cause significant response biases that influence HP in the partial design but not the complete design. To be clear, these biases related to congruency and which often interact with other factors such as alignment can only be measured in the complete design, but they can influence the partial design measure (even if d' is used) because congruency is confounded with correct responses.

However, that different stimulus manipulations influence response bias and the measure in the partial design but not the complete design does not tell us which measure best captures HP. On the one hand, the partial design measure could be indexing response biases that are caused by stimulus manipulations but are unrelated to HP (e.g., Cheung et al., 2008; Richler, Mack et al., 2011). On the other hand, it could be that these stimulus manipulations *should* affect HP, and these biases and HP are one and the same. For example, HP might be stronger for low spatial frequency faces *because* HP depends on coarse information that is extracted rapidly (e.g., Goffaux & Rossion, 2006); inversion might disrupt HP *because* the arrangement of parts no longer fits the face template (e.g., Rossion & Boremanse, 2008).

To determine whether the issue of response bias—which is already known to be influenced by alignment, congruency, and several other stimulus manipulations—invalidates the partial design operational definition of HP, we consider factors that are known to influence response bias but should theoretically not influence HP. It is generally agreed that HP is not under top-down control, either because it results from the operation of an encapsulated module that is not cognitively penetrable (Robbins & McKone, 2007) or because expertise has increased the automaticity of this strategy (Richler, Wong et al., 2011; Wong, Palmeri, & Gauthier, 2009). In fact, the very logic of the composite task assumes that HP can be measured despite instructions to selectively attend to face parts: The idea is to measure observers' automatic propensity to use more than the information to which they are instructed to selectively attend. Accordingly, a valid measure of HP should index the extent to which faces are processed holistically independently of participants' beliefs about the task, whether these are formed due to explicit instructions or are learned over the course of the experiment.

Here, we consider whether in addition to *stimulus* manipulations, a manipulation of *strategy* will differentially affect these operational definitions of HP within the composite paradigm. In the partial design, holistic processing is inferred from an alignment effect: Accuracy on "same" trials is higher when test composites are misaligned vs. aligned. In the complete design, holistic processing is indexed by a congruency $\times$ alignment interaction in d': Performance is better on congruent vs. incongruent trials, and the magnitude of this congruency effect is larger on aligned vs. misaligned trials. Thus, while both designs involve a manipulation of misalignment that reduces interference, only the complete design includes congruency as a factor (all "same" trials are incongruent in the partial design). Strategy was manipulated in two ways: In Experiment 1, we told participants that there would be mostly same or mostly different trials in the experiment, despite the fact that the true proportion of same/different trials was always 50%; in Experiment 2, we gave no explicit instructions about proportion, but the proportion of same and different trials was actually manipulated. We present evidence that the partial design index of HP is influenced by beliefs and strategy, which is inconsistent with the underlying theoretical construct it is presumed to index. However, the complete design provides a measure of HP that is robust to such influences.

Experiment 1

Methods

Participants

Thirty-six members of the Vanderbilt University community (18 males; mean age = 21.5 years) participated in exchange for $12. Participants were randomly assigned to the 75% same group (n = 18) or the 75% different group. The experiment was approved by the local Institutional Review Board.

Stimuli

Images of twenty female faces from the Max Planck Institute Database (Troje & Bulthoff, 1996) were converted

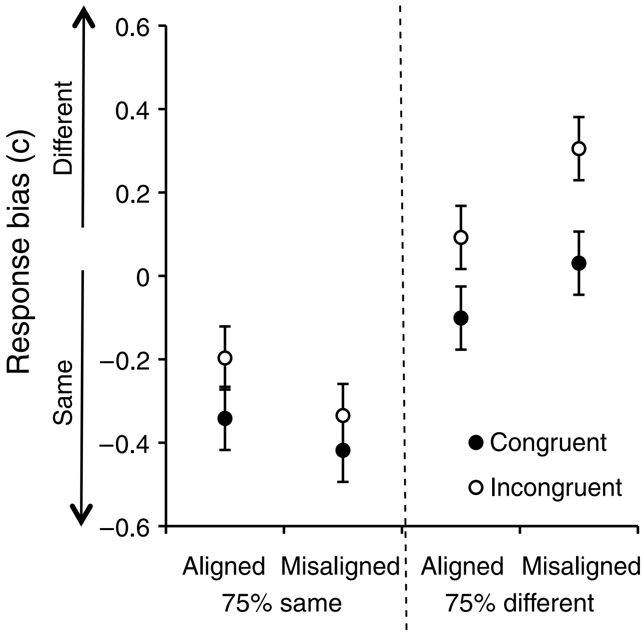

Figure 2. Response bias (*c*) on congruent and incongruent trials as a function of alignment for each instruction group. Positive values indicate a bias to respond "different" and negative values indicate a bias to respond "same." Error bars show 95% confidence intervals of the interaction.

to grayscale and cut in half to produce 20 face top halves and 20 face bottom halves, each 256 × 128 pixels in size. Face halves were randomly combined on every trial. A white line 3 pixels thick separated face halves resulting in faces that were 256 × 259 pixels. The white line was added to make it unambiguous where the top half ends and the bottom half begins, which, if anything, should facilitate selective attention to one-half. Misaligned faces were created by moving the top half of the face to the right by 35 pixels and the bottom half of the face to the left by 35 pixels, such that the edge of one face half fell in the center of the other face half.

Procedure

On each trial (see Figure 1, bottom panel), a fixation cross was presented (500 ms), followed by the study face (600 ms), and then a random pattern mask (500 ms). Next, a square bracket was presented (300 ms) that cued participants as to which part of the test face they would be judging. Then, the test face was presented. The cue remained on the screen during the presentation of the test face. Participants responded by key press if the cued part of the test face matched the corresponding part of the study face. The test face remained on the screen until participants responded or for a maximum of 2500 ms. Timeouts were extremely rare (<1% of trials) and these trials were not included in our analyses. The study face was always aligned. The test face could be either aligned or misaligned.

Beliefs were manipulated using instructions: The 75% same group was instructed that the target part would *remain the same* on 75% of the trials, while the 75% different group was told that it would *change* on 75% of the trials.

The complete design of the composite task (that contains the partial design trials, see Figure 1) was used. There were 10 trials for each combination of cued part (top/bottom), congruency (congruent/incongruent), alignment (aligned/misaligned), and correct response (same/different). The true proportion of same/different trials in the experimental block was always 50%, regardless of instruction condition. Therefore, the two groups only differed in their beliefs about the target part.

The experimental block was preceded by a 16-trial practice block, during which the proportion of same/different trials matched the instructions (i.e., 12/16 trials were "same" for the 75% same group).

Results

Top and bottom trials were combined in all analyses. Data from all trial types were included in the complete design analyses. Only data from trials where the irrelevant half was "different" (trials in gray boxes in Figure 1) were included in partial design analyses.

Analysis of response bias [c; $-0.5 * (\text{zHIT} + \text{zFA})$] calculated from the entire data set confirmed the effectiveness of our instructions. A 2 (congruency: congruent/incongruent) × 2 (alignment: aligned/misaligned) × 2 (instructions: 75% same/75% different) ANOVA on response bias revealed a main effect of instructions ($F_{1,34} = 18.19$, MSE = 0.324, $p < 0.0001$, $\eta_p^2 = 0.349$). As can been seen from Figure 2, the 75% same group was biased to respond "same," and the 75% different group was biased to respond "different." There was also a main effect of congruency ($F_{1,34} = 26.28$, MSE = 0.100, $p < 0.0001$, $\eta_p^2 = 0.436$) and an interaction between alignment and instructions ($F_{1,34} = 7.02$, MSE = 0.041, $p < 0.05$, $\eta_p^2 = 0.171$): Both groups were biased to respond "different" on incongruent trials, and participants in the 75% different group were also biased to respond "different" on misaligned trials (see Figure 2).

How did biases created by beliefs affect HP measures? As can be appreciated from Figure 3 (top left), the complete design measure was unaffected. A similar 2 × 2 × 2 ANOVA on sensitivity [d'; zHIT − zFA] in the complete design revealed evidence for HP by a significant congruency × alignment interaction ($F_{1,34} = 20.47$, MSE = 0.222, $p < 0.0001$, $\eta_p^2 = 0.376$). Critically, this interaction was not modulated by a three-way interaction with instructions ($F_{1,34} = 0.004$, MSE = 0.222, $p > 0.9$, $\eta_p^2 < 0.001$).

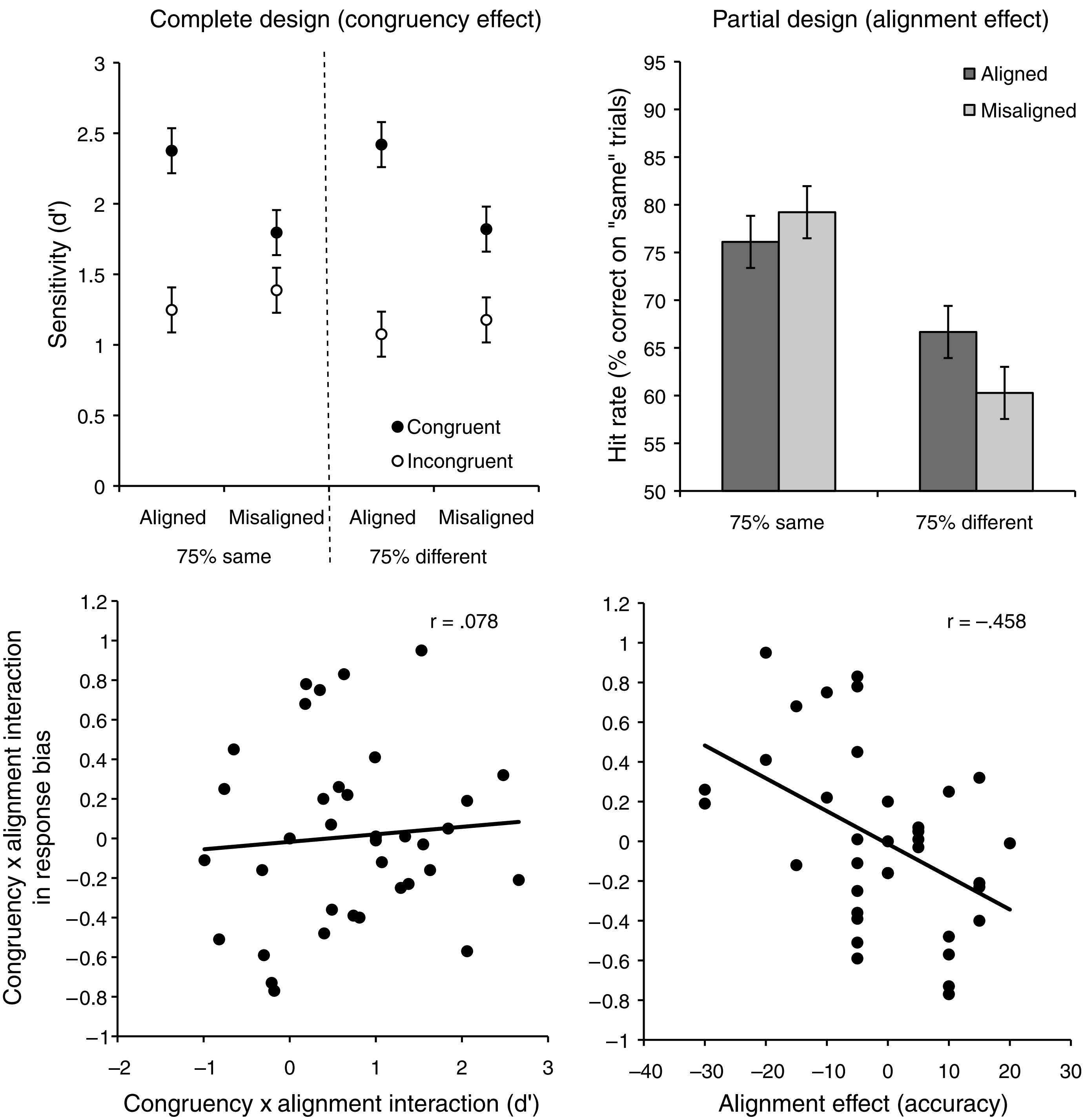

Figure 3. Top left panel: Sensitivity (d') on congruent and incongruent trials as a function of alignment for both instruction groups. Top right panel: Hit rate (percent correct on “same”-incongruent trials) as a function of alignment for each instruction group. Error bars show 95% confidence intervals of the interaction. Bottom left panel: Correlation between response bias (congruency × alignment interaction in c) and HP measured in the complete design (congruency × alignment interaction in d'). Bottom right panel: Correlation between response bias (congruency × alignment interaction in c) and HP measured in the partial design (alignment effect in hit rate).

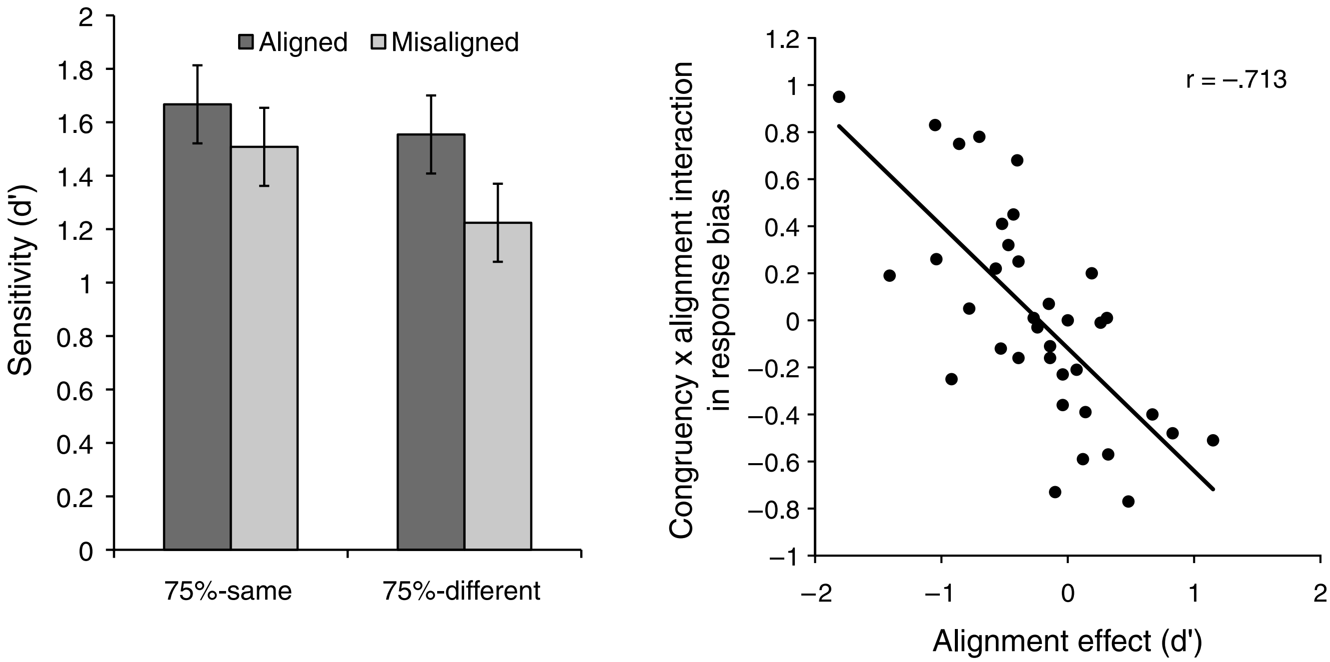

Figure 4. Left panel: Sensitivity (d') for partial design trials (same-incongruent and different-congruent) as a function of alignment. Error bars show 95% confidence intervals of the interaction. Right panel: Correlation between response bias calculated from the complete design (congruency × alignment interaction in c) and the alignment effect (misaligned minus aligned) in partial design d'.

In fact, no effect of instructions reached significance ($ps > 0.4$, $\eta_p^2 < 0.02$). Both groups performed better on aligned than on misaligned trials ($F_{1,34} = 10.63$, MSE = 0.187, $p < 0.01$, $\eta_p^2 = 0.238$) and on congruent than on incongruent trials ($F_{1,34} = 34.96$, MSE = 0.800, $p < 0.0001$, $\eta_p^2 = 0.507$).

Conversely, HP measured by the partial design (difference in accuracy on "same"-incongruent trials between misaligned and aligned trials) was influenced by our instructions. A 2 (alignment: aligned/misaligned) × 2 (instructions: 75% same/75% different) ANOVA on hit rate (percent correct on "same"-incongruent trials) revealed a main effect of instructions ($F_{1,34} = 8.18$, MSE = 459.11, $p < 0.01$, $\eta_p^2 = 0.194$), such that the 75% same group performed better overall. Critically, instructions interacted with alignment ($F_{1,34} = 6.91$, MSE = 65.16, $p = 0.013$, $\eta_p^2 = 0.169$). As can be seen in Figure 3 (top right), the alignment effect was not significant for the 75% same group ($t_{17} = 1.61$, $p > 0.1$, $d = 0.165$), but the *reverse* effect approached significance for the 75% different group ($t_{17} = 2.08$, $p = 0.053$, $d = 0.492$).

Importantly, the alignment effect in the partial design is related to the response biases created by our instructions. We calculated the magnitude of the congruency × alignment interaction in the complete design [(aligned-congruent minus aligned-incongruent) − (misaligned-congruent minus misaligned-incongruent)] in response bias (c)[2] and d', as well as the magnitude of the alignment effect (misaligned minus aligned) in hit rate (percentage of correct "same"-incongruent responses) for each participant. Correlational analyses showed that the magnitude of the congruency × alignment interaction in response bias calculated from the complete design predicted the magnitude of the alignment effect in the partial design ($r_{36} = -0.458$, $p < 0.01$, CI = 0.153, 0.683; see Figure 3, bottom right) but not the congruency × alignment interaction in d' in the complete design ($r_{36} = 0.078$, $p = 0.650$, CI = -0.257, 0.396; Figure 3, bottom left).

Although hit rate is the common measure in the partial design, a d' can be calculated for partial design trials (hereafter referred to as partial design d'; e.g., Konar et al., 2010). Although partial design d' is independent of response biases created by alignment, it is not independent from response biases related to congruency. Critically, even though congruency is not the manipulation of interest in the partial design, it is fully confounded with correct responses (all "same" trials are congruent, and all "different" trials are incongruent; see Figure 1). Therefore, the partial design d' cannot dissociate the contribution of response biases from the holistic effects that are both associated with congruency. Indeed, while a 2 × 2 ANOVA on partial design d' reveals a main effect of alignment ($F_{1,34} = 5.80$, MSE = 0.186, $p < 0.05$, $\eta_p^2 = 0.146$) that does not interact with instructions ($F_{1,34} = 0.710$, MSE = 0.186, $p = 0.405$, $\eta_p{}^2 = 0.020$), this effect is in the wrong direction: Performance is *better* for aligned vs. misaligned trials (Figure 4, left). Furthermore, the magnitude of the congruency × alignment interaction in response bias calculated from the complete design data predicts the magnitude of the alignment effect (misaligned minus aligned) in partial design d' (r_{36} = −0.713, p < 0.001, CI = −0.843, −0.503; Figure 4, right).

Discussion

In Experiment 1, we showed that manipulating participant strategy via instructions influenced HP measured in the partial design but not HP measured in the complete design. On the one hand, for those familiar with signal detection theory (Green & Swets, 1966), this is not surprising: We showed that hit rate is influenced by response bias and that d' is not. However, theoretically this is unexpected: If the partial design is truly capturing HP, then HP is dependent on participant beliefs and strategies, which is inconsistent with all views of HP, whether they posit that HP arises due to an automatic tendency to treat face parts non-independently (e.g., Richler et al., 2008) or due to the obligatory use of a face template during encoding (e.g., Rossion & Boremanse, 2008). Therefore, *why* the partial design measure is influenced by response bias is not surprising (this is after all the premise of signal detection theory), but *that* it happens at all is critical for the interpretation of this measure.

Understanding why partial design d' is also influenced by response bias is less intuitive. As described above, d' should be immune to influences of response bias. While this is true, in the partial design d' is calculated from two conditions: same-incongruent trials contribute to the hit rate, and different-congruent trials contribute to the false alarm rate. Because congruency is confounded ("same" trials are incongruent, and "different" trials are congruent), even though partial design d' is independent of biases related to alignment, it is impossible to remove the biases associated with congruency. The fact that the alignment effect measured in partial design d' is predicted by response bias based on congruency and alignment (as shown in Figure 4, right panel) indicates that these confounded biases are critical.

Still, it may be argued that we have manipulated strategy and bias in an artificial manner that does not correspond to typical experimental situations. In Experiment 2, we test whether strategy and response bias are influenced by an experimental design feature that is often used in this literature. Because only "same" trials are analyzed in the partial design, some authors include more "same" than "different" trials in their experiments to increase the amount of usable data (e.g., Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010; de Heering, Houthuys, & Rossion, 2007; Michel, Corneille, & Rossion, 2007, 2010; Ramon, Busigny, & Rossion, 2010; Rossion & Boremanse, 2008). In these experiments, the different trials are, in essence, catch trials. However, if explicit instructions about the proportion of same/different trials can influence strategy, response bias, and ultimately HP in the partial design, it is also possible that implicit learning of the proportion of same/different trials will also influence these factors. This possibility is tested in Experiment 2.

Experiment 2

Methods

Participants

Thirty-seven Vanderbilt University undergraduates (8 males; mean age = 20.12 years) participated in exchange for course credit. Participants were randomly assigned to the 75% same group (n = 19) or the 75% different group (n = 18). Data from one participant in the 75% same group was discarded because they used incorrect response keys. The experiment was approved by the local Institutional Review Board.

Stimuli and procedure

The stimuli and procedure were identical to Experiment 1 with the exception that for participants in the 75% same group, 75% of the trials were "same" trials, and for participants in the 75% different group, 75% of the trials were "different" trials. Regardless of the overall number of "same" or "different" trials, both trial types contained an equal proportion of congruent and incongruent trials. Importantly, no explicit instructions about the proportion of same/different trials were given. In addition, in Experiment 2, there were twice as many trials as Experiment 1 to increase reliability in the 25% trial conditions for each group (e.g., for the 75% same group, there were 240 "same" trials and 80 "different" trials).

Results

Response bias (c) based on congruency and alignment is shown in Figure 5. Unlike Experiment 1, our manipulation did not appear to systematically influence response bias. A 2 (alignment: aligned/misaligned) × 2 (congruency: congruent/incongruent) × 2 (proportion: 75% same/75% different) ANOVA on response bias only revealed a significant main effect of congruency, such that participants were more likely to respond "different" on incongruent trials ($F_{1,34}$ = 4.80, MSE = 0.043, p < 0.05, η_p^2 = 0.124). No effects of proportion reached significance (ps > 0.2, η_p^2 < 0.1).

However, did our manipulation of proportion influence HP? A similar 2 × 2 × 2 ANOVA on d' in the complete design (Figure 6, top left) revealed a significant main effect of congruency, such that performance was better on congruent vs. incongruent trials ($F_{1,34}$ = 67.74,

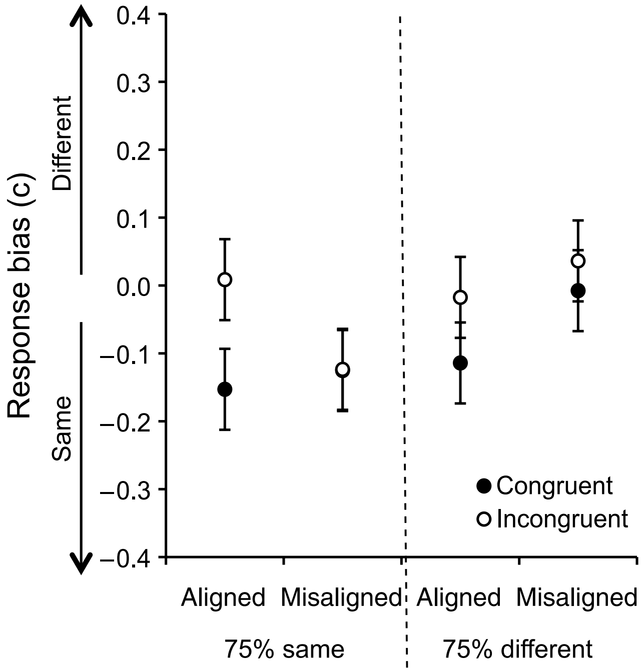

Figure 5. Response bias (*c*) on congruent and incongruent trials as a function of alignment for each instruction group. Note that the data points for misaligned-congruent and misaligned-incongruent for the 75% same group are superimposed. Positive values indicate a bias to respond "different" and negative values indicate a bias to respond "same." Error bars show 95% confidence intervals of the interaction.

MSE = 0.348, $p < 0.001$, $\eta_p^2 = 0.666$). Critically, although there was a significant main effect of proportion, with better overall performance when 75% of the trials were "same" ($F_{1,34} = 4.13$, MSE = 0.920, $p = 0.050$, $\eta_p^2 = 0.108$), both groups processed faces holistically, and HP did not vary based on proportion. There was a significant congruency × alignment interaction ($F_{1,34} = 22.32$, MSE = 0.195, $p < 0.001$, $\eta_p^2 = 0.396$) that was not modulated by an interaction with proportion ($F_{1,34} = 0.005$, MSE = 0.195, $p > 0.9$, $\eta_p^2 < 0.001$), nor did proportion interact with any other factor ($ps > 0.1$, $\eta_p^2 < 0.1$).

In contrast, we found at least some evidence that the partial design measure of HP was not so robust to this manipulation of proportion. A 2 (alignment: aligned/misaligned) × 2 (proportion: 75% same/75% different) ANOVA was conducted on hit rate (accuracy on "same"-incongruent trials). As can be appreciated from Figure 6 (top right), although the interaction between alignment and proportion did not reach statistical significance ($F_{134} = 2.64$, MSE = 134.52, $p = 0.113$, $\eta_p^2 = 0.072$), the alignment effect was significant when 75% of trials were "same" ($t_{17} = 3.86$, $p < 0.01$, $d = 1.32$) but not when 75% of trials were "different" ($t_{17} = 0.165$, $p = 0.871$, $d = 0.057$).[3] Moreover, although we did not observe significant group-level effects in response bias, response bias calculated in the complete design predicted the alignment effect in the partial design ($r_{36} = -0.628$, $p < 0.001$, CI = −0.378, −0.792; Figure 6, bottom right) but not the congruency × alignment interaction in d' in the complete design ($r_{36} = -0.257$, $p = 0.130$, CI = −0.539, 0.078; Figure 6, bottom left).

A similar 2 × 2 ANOVA on partial design d' (Figure 7, left) revealed a trend toward better overall performance when 75% of trials were "same" ($F_{1,34} = 3.58$, MSE = 0.494, $p = 0.067$, $\eta_p^2 = 0.095$). There were no other significant effects ($ps > 0.1$, $\eta_p^2 < 0.1$) nor was the alignment effect significant for either group ($ps > 0.2$, $ds < 0.3$). Response bias (based on alignment and congruency calculated in the complete design) predicted the magnitude of the alignment effect in partial design d' ($r_{36} = -0.772$, $p < 0.001$, CI = −0.877, −0.595; Figure 7, right).

Discussion

In Experiment 2, we showed that manipulating the proportion of same/different trials in the composite task influences whether or not evidence for HP is observed in the partial design, such that no evidence for HP was observed when 75% of trials were "different."[4] In other words, the experimental context (e.g., viewing "same" trials in the context of more "different" trials) can influence whether or not HP is observed in the partial design. This is troubling, since no theory of HP for faces would expect it to be dependent on proportion of trials. In contrast, manipulating proportion of same/different trials did not influence HP measured in the complete design, consistent with previous work showing that contextual manipulations can influence failures of selective attention in novices, but that HP in experts is much more robust to such influences (see Richler, Wong et al., 2011 for a review).

General discussion

Stimulus manipulations can introduce response biases that impact the alignment effect measured in the partial design (Cheung et al., 2008; Richler, Mack et al., 2011). Here, we show that this measure of HP is also susceptible to *strategic* manipulations of response bias, created either through explicit instructions (Experiment 1) or learning over the course of the experiment (Experiment 2). It is possible that physical differences in the stimuli (e.g., orientation, alignment, spatial frequency content) sometimes induce similar strategic responses, explaining why response bias can vary across conditions within an experiment. In addition, our results confirm prior reports for a considerable amount of variability in subjects' biases in this task (e.g., Richler, Cheung et al., 2011; Richler,

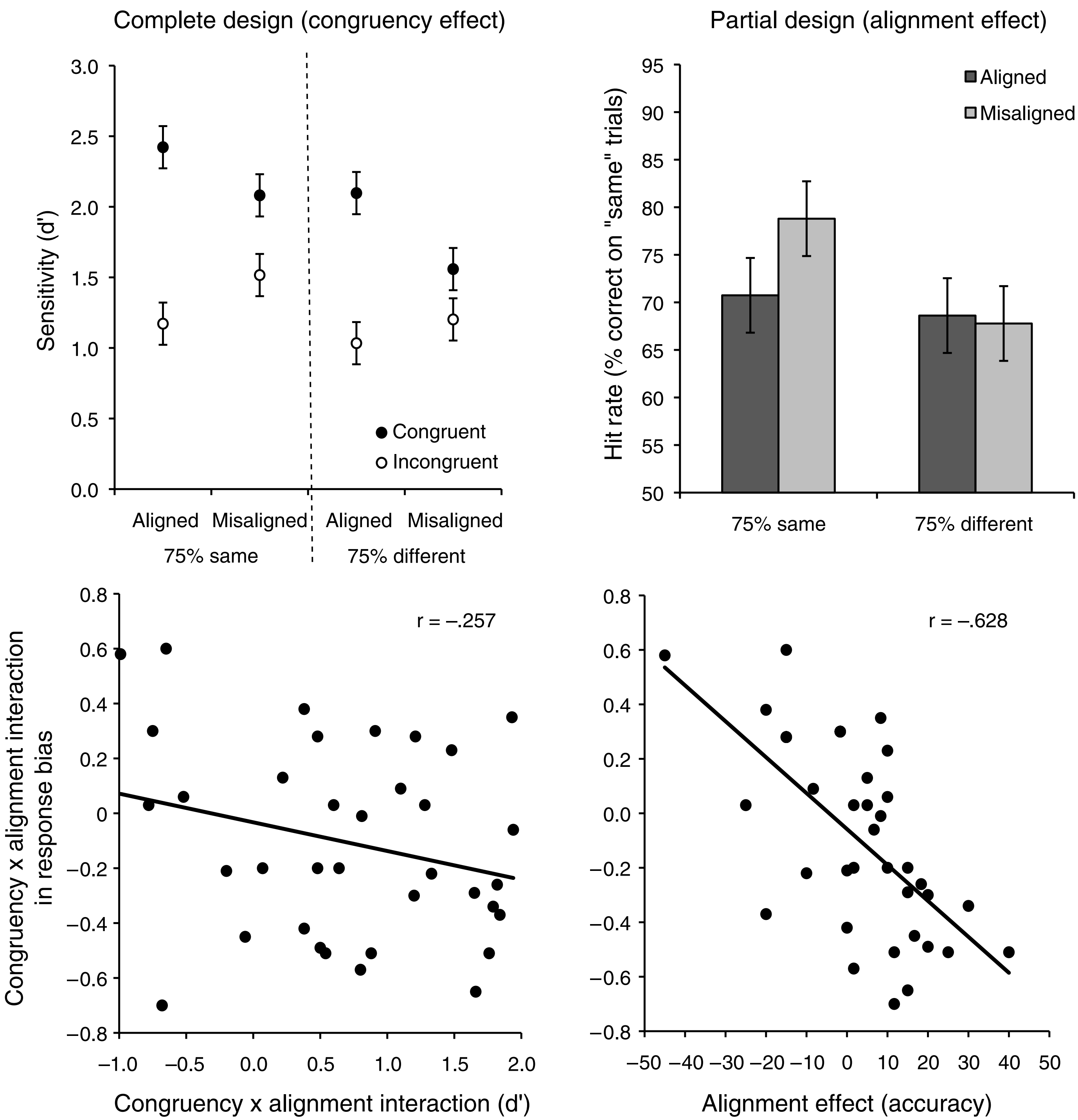

Figure 6. Top left panel: Sensitivity (d') on congruent and incongruent trials as a function of alignment for both instruction groups. Top right panel: Hit rate (percent correct on "same"-incongruent trials) as a function of alignment for each instruction group. Error bars show 95% confidence intervals of the interaction. Bottom left panel: Correlation between response bias (congruency $\times$ alignment interaction in c) and HP measured in the complete design (congruency $\times$ alignment interaction in d'). Bottom right panel: Correlation between response bias calculated from the complete design (congruency $\times$ alignment interaction in c) and HP measured in the partial design (alignment effect in hit rate).

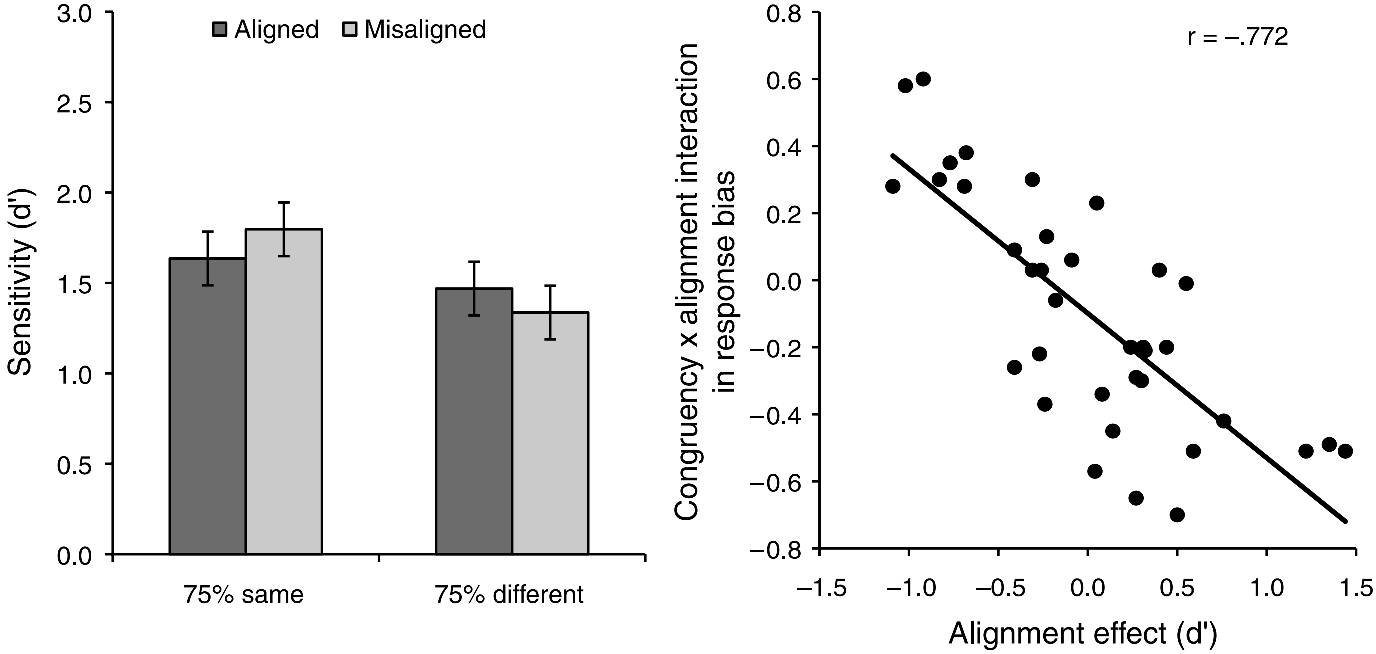

Figure 7. Left panel: Sensitivity (d') for partial design trials (same-incongruent and different-congruent) as a function of alignment. Error bars show 95% confidence intervals of the interaction. Right panel: Correlation between response bias calculated in the complete design (congruency $\times$ alignment interaction in c) and the alignment effect (misaligned minus aligned) in partial design d'.

Mack et al., 2011). While it is unknown whether these individual differences in bias are stable, within the course of an experiment, they are large enough to contribute about 60% of the variance in the partial measure of HP. The problem is not that the partial design measure is not affected by HP but that it is also sensitive to such biases.

In one sense, our results are not surprising: We showed that manipulating response bias influences hit rate but not d', which is the very premise of signal detection theory (Green & Swets, 1966). However, the demonstration that biases—especially strategically driven biases—impact both hit rate and d' measures of HP in the partial design is far from trivial for the field of face processing. Perhaps most importantly, this suggests that the partial design at best measures HP as a state but perhaps less as a stable trait or a skill. Indeed, if beliefs can influence HP through effects on response bias, then other factors that can also influence response bias can contribute to individual differences in HP. This may explain why recent individual difference studies showed no relationship between the partial design measure of HP and face recognition ability (Konar et al., 2010; Richler, Cheung et al., 2011), while the complete design measure of HP correlated with face recognition (Richler, Cheung et al., 2011).

These findings have important theoretical implications because of the pervasiveness of the partial design; for instance, while young children (e.g., Cassia, Picozzi, Kuefner, Bricolo, & Turati, 2009) or patients with schizophrenia (Schwartz, Marvel, Drapalski, Rosse, & Deutsch, 2002) may indeed exhibit normal, adult-like HP, these results could also be artifacts of important (and potentially informative) group differences in response biases, as these claims are only supported by partial design experiments. For example, it is well known that young children exhibit a bias to respond affirmatively (e.g., Fritzley & Lee, 2003; Okanda & Itakura, 2010). This affirmative bias could also influence the partial design measure of HP. More generally, using a measure that confounds response bias and sensitivity inflates measurement error in assessments of HP, with all the associated negative consequences (reduced power, unstable effect size estimates, increased Type II errors in individual studies, increased Type I errors in the resulting body of research; see Ellis, 2010).

Because the complete and partial versions of the composite task often yield qualitatively incompatible results, it is necessary to compare how they fare in accounting for the existing body of evidence on HP. To keep using these two measures for the same construct is not an option since they will not converge, and they should not be combined in informal or formal meta-analytical reviews on this topic. To validate measures of unobservable concepts such as HP, it is critical to evaluate their role in nomological networks that relate observable properties to unobservable constructs and different theoretical constructs to each other (Cronbach & Meehl, 1955).

We contend that the partial design measure of HP leads to a weak nomological network. First, using this measure would force us to accept that HP is not related to face recognition ability (Konar et al., 2010; Richler, Cheung et al., 2011), which is incompatible with the current status of HP as a hallmark of face perception. Second, some studies based on the partial design find that HP is not observed in perceptual experts for objects (Robbins & McKone, 2007), but the theoretical importance of that finding is unclear if HP is not related to expertise for faces in the first place. In addition, theories of HP suggest that HP should not be under top-down control (Robbins & McKone, 2007; Wong, Palmeri, & Gauthier, 2009), yet here we show that the partial design measure of HP is influenced by participants' beliefs and strategies. In sum, this methodological issue can lead to theoretical confusion. For instance, it is paradoxical that authors who argue that HP is perceptual (Jacques & Rossion, 2009; Kuefner, Jacques, Prieto, & Rossion, 2010) have mainly relied on a measure of HP that is so sensitive to top-down influences.

In contrast, when HP is measured in the complete design, results are consistent with theoretical constructs: HP is obtained for upright faces for stimulus presentations as rapid as 50 ms (Richler, Mack et al., 2009), while longer exposure durations are required for HP to be obtained with inverted faces (Richler, Mack et al., 2011), and HP is not obtained with non-face objects in novices (Richler, Bukach, & Gauthier, 2009; Richler, Mack et al., 2011). HP relates to expertise for both faces (Richler, Cheung et al., 2011) and objects (Gauthier & Tarr, 2002; Wong, Palmeri, & Gauthier, 2009) and correlates with face-specific neural markers (e.g., right fusiform activity, Wong, Palmeri, Rogers, Gore, & Gauthier, 2009; N170 ERP component, Gauthier, Curran, Curby, & Collins, 2003). In addition, the complete design measure of HP is not influenced by stimulus-driven response biases (Cheung et al., 2008; Richler, Mack et al., 2009, 2011), consistent with most models of face perception that assume that HP is a perceptual effect (e.g., Turk & Pentland, 1991), nor is it influenced by response biases that reflect top-down strategies, as shown here.

Of course, HP can be measured in other tasks that might not be complicated by issues having to do with response bias. However, the complete composite paradigm provides a measure of HP that is robust to stimulus, task, and strategic manipulations that influence response bias (Cheung et al., 2008; Richler, Mack et al., 2009, 2011) while also measuring decisional effects that may be interesting in their own right. Other measures of HP may eventually compare favorably to the complete design of the composite task but each measure needs to be adequately validated. A measure of HP should at least (i) capture larger HP for faces than objects in novices, (ii) produce HP effects correlated with face recognition ability, and (iii) be robust to top-down manipulations of response bias. This is not an exhaustive list: Nomological networks grow with a literature and the best measure at any point in time is the one that supports the most coherent network.

Unfortunately, it is not uncommon for authors to adopt a measure based only on the first of the three aforementioned criteria, coupled with some degree of face validity and/or the fact that the measure has been used in the literature. Perhaps most disturbing for the growth of the field, some authors go as far as to reject the complete design because it leads to results that are inconsistent with what is "known" based on the partial design and verbal descriptions of the mechanisms underlying the composite illusion that are not supported empirically. For example, a recent paper (Palermo et al., 2011) suggested that the complete design is invalid because it leads to the finding that inverted faces are processed holistically (Richler, Mack et al., 2011). However, the absence of holistic processing for inverted faces is only supported by partial design data (e.g., Rossion & Boremanse, 2008) and has been shown to be related to response biases (Richler, Mack et al., 2011). It may be surprising that inverted faces are processed holistically, but Richler et al. also found that holistic processing of inverted faces has a protracted time course relative to upright faces. In other words, there are differences in how we process upright and inverted faces; this difference is just not as qualitative as is sometimes argued (see also Sekuler, Gaspar, Gold, & Bennett, 2004). In the light of our present argument, instead of viewing surprising and unexpected findings as a basis for rejecting the complete design, we should be pursuing these interesting results. Thus, although abandoning the partial design means that some phenomena need to be reinvestigated, the end result will be less inconsistencies in the field of face recognition, and the stage will be set to move forward and advance our understanding of what it means for faces to be processed holistically.

Acknowledgments

This work was supported by grants from the JSMF and NSF (SBE-0542013), the Vanderbilt Vision Research Center (P30-EY008126), and the National Eye Institute (2 R01 EY013441-06A2). We thank Magen Speegle, Yoseph Beki, and Alex Desire for data collection.

Commercial relationships: none.
Corresponding author: Jennifer J. Richler.
Email: jennifer.j.richler@vanderbilt.edu.
Address: Department of Psychology, Vanderbilt University, 111 21st Avenue South, Wilson Hall, Nashville, TN 37240, USA.

Footnotes

[1]We do not yet have a theory to predict the various response biases that occur in this task. The bias to respond "different" on incongruent trials is often found for faces (e.g., Cheung et al., 2008; Richler, Mack et al., 2009,

2011; Richler et al., 2008), but it has also been observed for novel objects in novices (Richler, Mack et al., 2011), suggesting a domain-general effect where congruency influences response bias even for categories that are not processed holistically. One possibility is that, even when a subject's ability to match the relevant part (d') is not affected by the irrelevant part, the identity of the irrelevant part is, to some extent, processed, and on incongruent trials, a conflict signal might sometimes be misinterpreted as a sign that the relevant face part has changed.

[2]We used the congruency × alignment interaction to index response bias because this provides a single measure that takes into account both factors and differences between conditions.

[3]An independent-samples *t*-test comparing the magnitude of the alignment effect between proportion groups revealed a Cohen's *d* of 0.557, indicating a medium–large effect size. In contrast, a similar *t*-test comparing the complete design HP measure between proportion groups revealed a Cohen's *d* of 0.024, indicating a negligible effect.

[4]Note that regardless of the proportion of same/different trials, the alignment effect was always calculated based on accuracy for "same" trials. Thus, not finding an effect in the 75% different condition is not the result of the composite illusion in the partial design only occurring on "same" trials (McKone & Robbins, 2007).

References

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia, 48,* 4057–4092. [PubMed]

Cassia, V. M., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science, 12,* 236–248. [PubMed]

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. [PubMed]

de Heering, A., Houthuys, S., & Rossion, R. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96,* 57–70. [PubMed]

Ellis, P. D. (2010). *The essential guide to effect sizes: An introduction to statistical power, meta-analysis and the interpretation of research results*. United Kingdom: Cambridge University Press.

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

Fritzley, V. H., & Lee, K. (2003). Do young children always say yes to yes–no questions? A metadevelopmental study of the affirmation bias. *Child Development, 74,* 1297–1313. [PubMed]

Gauthier, I., & Bukach, C. M. (2007). Should we reject the expertise hypothesis? *Cognition, 103,* 322–330. [PubMed]

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6,* 428–432. [PubMed]

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–466. [PubMed]

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception & Performance, 32,* 1023–1039. [PubMed]

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision, 9*(6):8, 1–16, http://www.journalofvision.org/content/9/6/8, doi:10.1167/9.6.8. [PubMed] [Article]

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

Kuefner, D., Jacques, C., Prieto, E. A., & Rossion, B. (2010). Electrophysiological correlates of the composite face illusion: Disentangling perceptual and decisional components of holistic face processing in the human brain. *Brain and Cognition, 74,* 225–238. [PubMed]

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15,* 762–768. [PubMed]

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11,* 8–15. [PubMed]

McKone, E., & Robbins, R. (2007). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition, 103,* 331–336. [PubMed]

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science, 31,* 911–924. [PubMed]

Michel, C., Corneille, O., & Rossion, B. (2010). Holistic face encoding is modulated by perceived face race: Evidence from perceptual adaptation. *Visual Cognition, 18,* 434–455.

Okanda, M., & Itakura, S. (2010). When do children exhibit a "yes" bias? *Child Development, 81,* 568–580. [PubMed]

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia, 49,* 1226–1235. [PubMed]

Ramon, M., Busigny, T., & Rossion, B. (2010). Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. *Neuropsychologia, 48,* 933–944. [PubMed]

Richler, J. J., Bukach, C. M., & Gauthier, I. (2009). Context influences holist processing of nonface objects in the composite task. *Attention, Perception & Psychophysics, 71,* 530–540. [PubMed]

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science, 20,* 129–134. [PubMed]

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009) Holistic processing of faces happens at a glance. *Vision Research, 49,* 2856–2861. [PubMed]

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

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory and Cognition, 34,* 1356–1368. [PubMed]

Richler, J. J., Wong, Y. K., & Gauthier, I. (2011). Perceptual expertise as a shift from strategic interference to automatic holistic processing. *Current Directions in Psychological Science, 20,* 129–134. [PubMed]

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioral tasks. *Cognition, 103,* 34–79. [PubMed]

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, http://www.journalofvision.org/content/8/4/3, doi:10.1167/8.4.3. [PubMed] [Article]

Schwartz, B. L., Marvel, C. L., Drapalski, A., Rosse, R. B., & Deutsch, S. I. (2002). Configural processing in face recognition in schizophrenia. *Cognitive Neuropsychiatry, 7,* 15–39. [PubMed]

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,* 391–396. [PubMed]

Troje, N., & Bulthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36,* 1761–1771. [PubMed]

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.

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

Wong, A. C.-N., Palmeri, T. J., Rogers, B. T., Gore, J. C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS ONE, 4,* e8405. [Article]

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