Monocular Facilitation Implicates Subcortical Involvement in Holistic Processing of Faces

Rebeka C. Almasi and Marlene Behrmann

Abstract

Holistic processing (HP) refers to obligatory processing of the entirety of a visual stimulus rather than independent parts. In particular, HP has been implicated in face processing. Given that HP of faces has evolutionary significance, and subcortical structures generally being evolutionarily older than neocortex, the present research tests whether subcortical structures might contribute to the composite face effect (CFE), a common indicator of HP. According to the CFE, identical top halves of faces are more likely to be judged as different when the two bottom halves are from different faces, and differing top halves are less likely to be noticed as different given identical bottom halves (and vice versa for judging the bottom). Using a mirror stereoscope, we tested whether the CFE is greater when the images are presented to a single eye (monocularly) as opposed to different eyes (binocularly) sequentially, which would indicate HP outside cortical involvement. This study contributes to the literature on subcortical mechanisms in face processing, and, ultimately, has implications for our understanding of the evolutionary history of face processing.

Introduction

Social interaction depends on identifying individuals, usually by their faces—from finding a friend's face in a crowd to recognizing parents, face processing is a crucial function for which humans tend to build expertise. With both imaging and purely behavioral studies, visual perception research continues to explore face processing as implicating a domain-specific or a potential separate cognitive module. Infants' attention is automatically drawn to faces over other stimuli, and faces are subject to inversion effects (identification is far more impaired when a face is presented upside down compared to other stimuli like houses) (Farah 1995; Johnson, Dziurawiec, Ellis, Morton 1991). The fusiform gyrus of the brain has areas that seem to preferentially activate for faces, although this region has also been shown to activate for objects in which the observer has expertise, suggesting that it may be more of an expertise or holistic processing (HP) area (Kanwisher 1997; Gauthier, Behrmann & Tarr 1999). The magnitude of holistic processing for nonface objects is correlated with FFA activity for those objects following individuation training (Richler and Gauthier 2014); HP is a potential explanation for why faces and certain objects display different patterns of activation and special behavioral results.

Holistic Processing

HP suggests that the response to the whole cannot be predicted from the sum of responses to the parts. Humans have evolved to recognize faces quickly and accurately. HP is often tested through the composite face effect paradigm (CFE), which manipulates faces by changing the bottom and/or top and occasionally offsetting top from bottom. The CFE means that identical top halves of faces are more likely judged as different when the two bottom-half faces are from

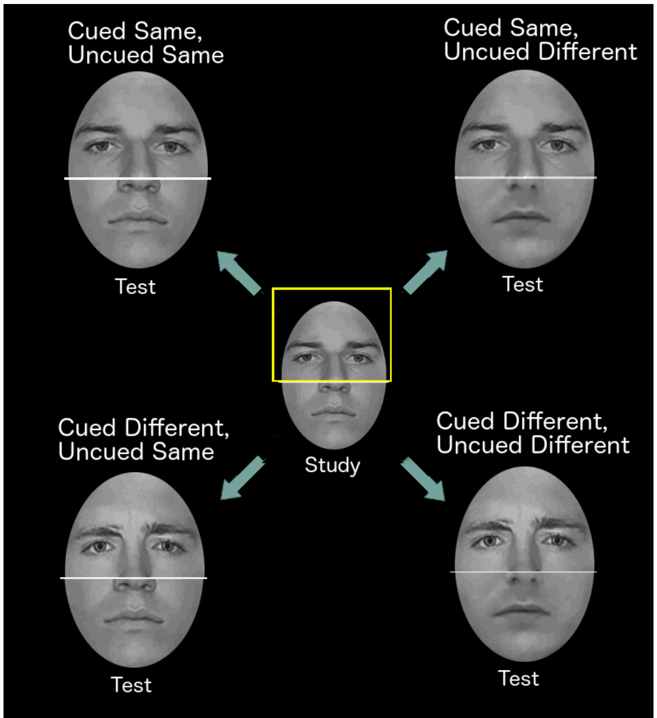


Figure 1 (above): Visualization of different composite faces; the "study" face in the center is displayed with one half highlighted, and the participant decides whether the "test" face displayed afterwards is the same person. The top left and bottom right are "congruent" trials because both halves change or do not change, while the top right and bottom left are incongruent and therefore more difficult to judge due to interference by HP.

different faces. Participants are instructed to attend to one half of the face and compare it to another face presented directly after, determining whether it is the same person or different (see Figure 1). Ultimately, the CFE tests failures of selective attention: participants cannot ignore the information in the irrelevant face half. When the face halves are misaligned, the meaningful face configuration is disrupted, reducing or eliminating holistic processing, as humans do not build up expertise with misaligned faces. When both the top and bottom either change or remain the same, the trial is said to be "congruent", but if only one changes, the trial is "incongruent"; it is these incongruent trials that require the selective attention that HP inhibits. Therefore, the expected results are poor performance on aligned incongruent trials which improves when the faces are misaligned. Thus, the CFE shows stronger effects of HP given aligned faces compared to misaligned.

Subcortical Involvement

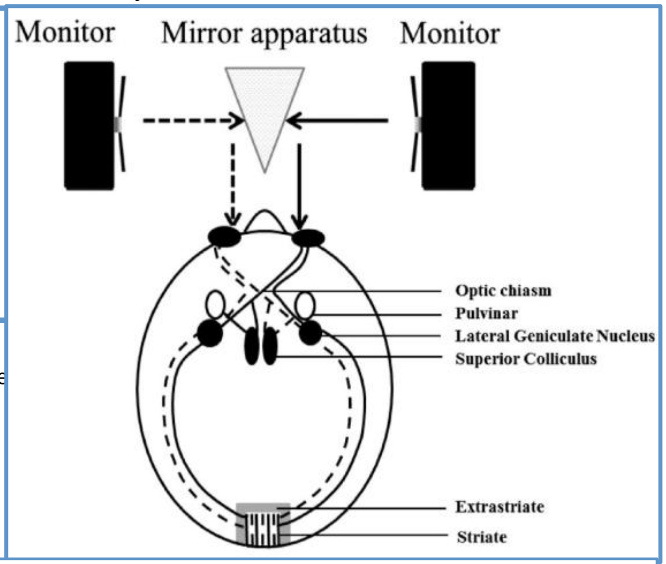
While most research on face perception focuses on a distributed cortical network of faces, recent research has suggested subcortical involvement as well. Evidence for subcortical processing includes imaging research that shows subcortical facilitation of objects for gray but not purple (which cannot be processed by the superior colliculus, a subcortical structure and studies suggesting that lower order species with primitive brain structures, such as paper wasps, animals without cortices can do complex visual perception tasks (Tamietto, 2010; Lin & Chiao, 2017). The amygdala has cells that respond selectively to faces, although they also respond to the names of the individuals pictured (Cauchoix & Crouzet, 2013). Finally, a monocular advantage for face perception compared to other objects is observed, even when the faces differed in location and size, implicating pre-cortical contribution to the face processing (Gabay, Nestor, Dundas, & Behrmann, 2014).

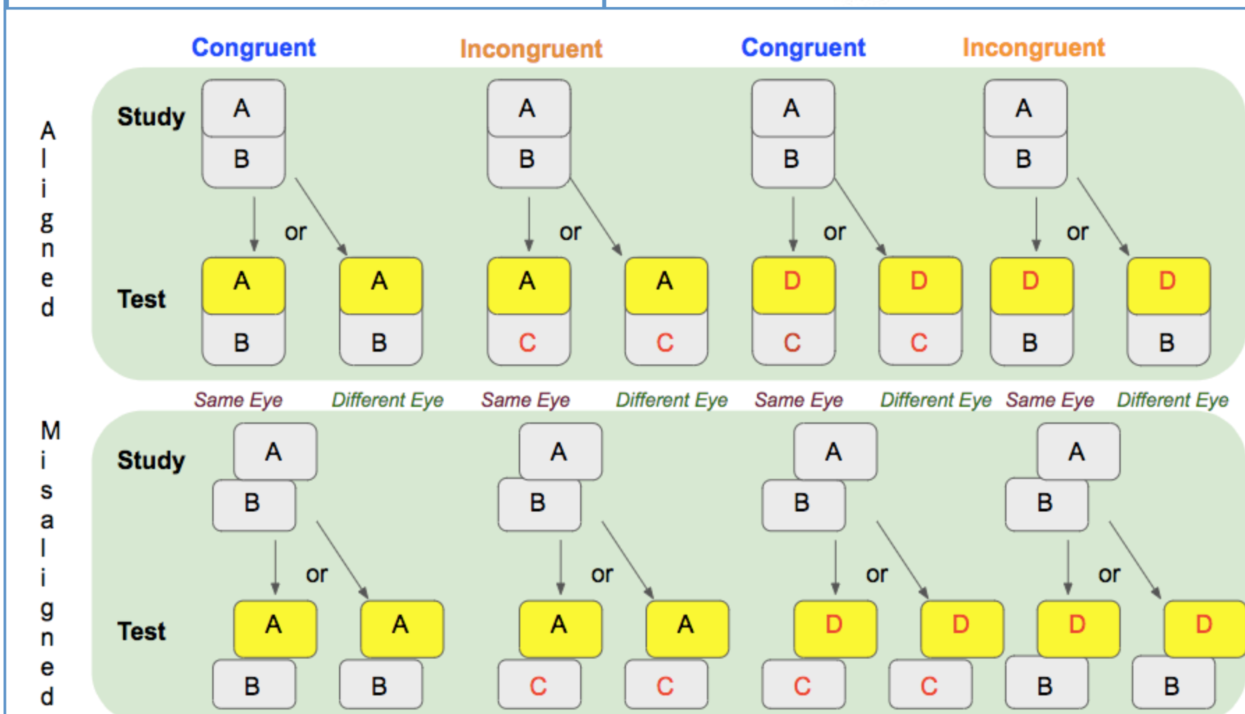
This "monocular advantage" is most relevant to the present study, as we use this same paradigm to circumvent the problem of subcortical brain imaging: the small size of the structures and their location deep in the brain lead to limited image resolution. Gabay et al., and the present study, used a mirror stereoscope to evaluate whether subcortical structures may play a role in pattern recognition. This technique takes advantage of the fact that visual input, once received by the retina, is propagated in an eye-specific fashion through the early stages of the visual system up until layer IV of V1 (Menon, Ogawa, Strupp, & Uğurbil, 1997) (see Figure 2). By presenting stimuli to only one eye, we can examine the involvement of monocular channels, which are mostly subcortical. If we find that presenting the same eye with two different images sequentially as opposed to different eyes contributes to CFE more, then this CFE (and therefore HP) is facilitated within low levels of the visual pathway.

The present study ties together two previous lines of research--the one into subcortical mechanisms of perception and the one into mechanisms of face processing. It tests whether participants show a monocular advantage for congruent conditions of composite faces as opposed to incongruent. Broadly, "How advanced does a brain have to be to be an expert in face processing?". We expected to find some monocular facilitation, which would imply much older mechanisms for HP than the cortical networks currently known.

Figure 2 (right): From Gabay, Nestor, Dundas, & Behrmann 2014: schematic representing the mirror stereoscope and segregation of visual pathways. A chin rest stabilizes the participant's head, and the mirrors are at 45 and 135 degrees reflecting two computer monitors 50 cm on the left and right side of the observers. Cardboard dividers blocked the participant's direct view of the monitors. The participants "fuse" the image by moving two fixation crosses, resulting in unawareness of which eye is receiving the stimulus.

Figure 3 (below): Visualization of conditions. For simplicity's sake, this diagram only displays trials where the "top" half of the face is cued rather the bottom, indicated by the yellow rectangle. The "test" face, which the participant judges to be same or different based on the cued half, is presented to either the same eye as the study face or the other eye.





Methods

Participants

26 participants recruited from the Psychology student pool at Carnegie Mellon University in Pittsburgh, PA with normal or corrected-to-normal vision were tested in total, of which 14 were female, 14 were Asian, 3 were Black, 8 were White, and 1 "other". One participant identified as Hispanic/Latino. The mean age was 19.96. Five participants were excluded based on left-handedness or inability to properly fuse images. All signed consent forms and were compensated with course credit for participation.

Stimuli

Face stimuli were created using 40 front-view Caucasian male faces from the Tanaka Lab face database. Each face was approximately 170 pixels in width and 240 pixels in height on a 320 × 420 pixel black background. Twenty faces were subdivided into five groups of four similar faces based on prior ratings (Liu & Behrmann 2014). Each composite face was created by pairing the top half of one face with the bottom half of another face from the same group. For misaligned trials, there was an approximately 80 pixel offset between the top and bottom half.

We used a complete composite design, comprising four different configurations: same-congruent (attended and unattended half both the same), different-congruent (attended and unattended half both different), same-incongruent (attended half same unattended different), and different-incongruent (attended half different with other half same) (see Figure 3). A complete composite design is preferred for HP experiments over a partial design (congruent different, incongruent same). In a partial design, the target face half is either the same as or different from the corresponding part of the study face, but the irrelevant halves are always different. This causes correctness to be confounded with congruency (Richler & Gauthier 2014). Therefore, the present study utilizes a complete composite design.

Procedure

Before each block, participants fused the images in the mirror stereoscope by moving the locations of two crosses left or right (one on each screen) using the keyboard until they overlapped. These locations would then be where stimuli and fixations appeared throughout the experiment. Each participant completed 4 blocks of 200 trials each, totaling 800 trials. Two of the blocks were all misaligned and two consisted purely of aligned stimuli. The ordering was counterbalanced between subjects. Trial types were otherwise randomly interleaved. Participants fused images again before each block to ensure head movement between blocks would not affect fusion, and rests of at least ten seconds were forced every 25 trials. Half of the trials were presented to the same eye sequentially for both test and study, evenly distributed between left and right, and the other half was evenly left-study-right-test and vice versa--this eye segregation was achieved through use of a mirror stereoscope (see Figure 4). This is a important distinction:

in this paper, “binocular” refers to presenting to different eyes, not to both eyes at once. Participants were instructed to attend to either the top or bottom of each face, indicated by a yellow rectangular cue framing the cued half. Participants judged whether the attended halves of the two faces were the same or not by pressing a corresponding key on a keyboard--either “f” or “j” referred to “same” and the other to “different”, with the assignments counterbalanced between participants. Accuracy was used to calculate d’ through aggregating the matlab data, meant to determine the participant’s discrimination sensitivity to measure holistic processing.

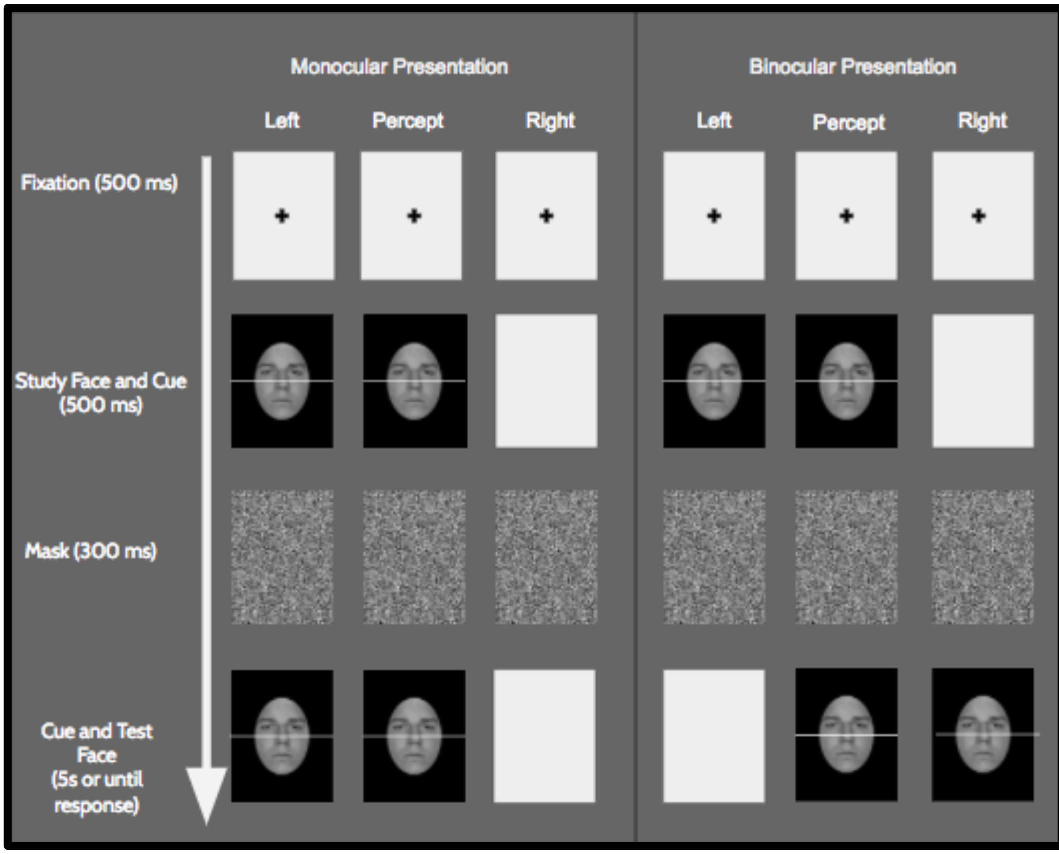


Figure 4: Visualization of appearance of screens. Each trial began with a black fixation cross presented at the center of the gray screen for 500 ms. After that, a study composite face was shown for 500 ms, together with a square bracket cueing which half of the face (left or right half) was to be judged. The study face was followed by a 300-ms mask. A test composite face was then displayed for 5 s or until a response was made. Note that this figure only portrays trials that begin with presentation on the left, which is only half.

Statistical methods

Matlab was used to collect the data, while data analysis was performed in Excel and the statistical software R, version 3.1.2. Data were cleaned as follows: first, distributions of correct answers were inspected and blocks with accuracy levels below 75% as well as individual trials with reaction times over 3 seconds were excluded entirely. The first 5 trials of each block were also excluded from further analysis. Percent accuracy was calculated and reaction time was aggregated within subject over all remaining trials for each combination of the four conditions, namely: monocular/binocular condition, cued side, alignment, and congruent/incongruent cues, and used in all subsequent analyses. D’, or discrimination sensitivity, values were calculated based on the accuracy and false alarm rates within each condition. The distribution of d’ values and reaction times was graphically inspected using boxplots. Even after deletion of below-75%

blocks and first 5 trials, several outliers were detected; these were winsorized, i.e. censored by replacing them with the nearest non-outlier value. We modeled each of the two measures separately, using mixed effect modeling. This technique can be applied to repeated measures data, including within-subject designs with missing values for certain subject/condition combinations, as long as the data is missing at random. ANOVA was used to analyze the four condition variables (Monocular/Binocular, Alignment, Cued Side (Top/Bottom), Congruency) and all their interactions, including the 4-way interaction. Within-subject correlations were modeled by subject-specific random intercepts. Finally, after seeing the cued side contribution, we performed a post-hoc paired t-test examining the d' differences between aligned and misaligned trials to better examine the difference between bottom and

Results

Data from 21 participants were eventually included in the final dataset. Average d' over all condition combinations was 2.25±0.91. To suggest monocular pathway involvement in HP, the results would predict a higher magnitude of composite face effect for monocular presentation: this difference between monocular and binocular should be most obvious for incongruent, aligned trials. While we did not find this exact pattern, we did find interactions replicating the composite face effect with ANOVA.

Term	F-value	P-value
Eye	0.00	0.9751
Cued Side	54.54	**<.0001**
Congruency	44.75	**<.0001**
Alignment	0.03	0.8579
Eye * Cued Side	2.43	0.1205
Eye * Congruency	5.36	**0.0213**
Cued Side * Congruency	2.78	0.0964
Eye * Alignment	0.03	0.8725
Cued Side * Alignment	0.67	0.4123
Congruency * Alignment	14.94	**0.0001**
Eye * Cued Side * Congruency	0.11	0.7396
Eye * Cued Side * Alignment	3.61	0.0585
Eye * Congruency * Alignment	0.68	0.4112
Cued Side * Congruency * Alignment	6.32	**0.0125**
Eye * Cued Side * Congruency * Alignment	3.33	0.0691

Table 1: ANOVA results. The congruency * alignment interaction is later discussed as part of the 3-way interaction. Cued Side * Congruency * Alignment. While the four-way interaction is not below the commonly accepted .05 threshold in significance, the strength of the trend still warrants exploration.

In the mixed effect model with all four conditions, all possible interactions as independent variables, the highest-order interaction neared significance (4-way interaction F=3.33,1, df=(1, 299), p=0.0691) (Figure 5a,b).

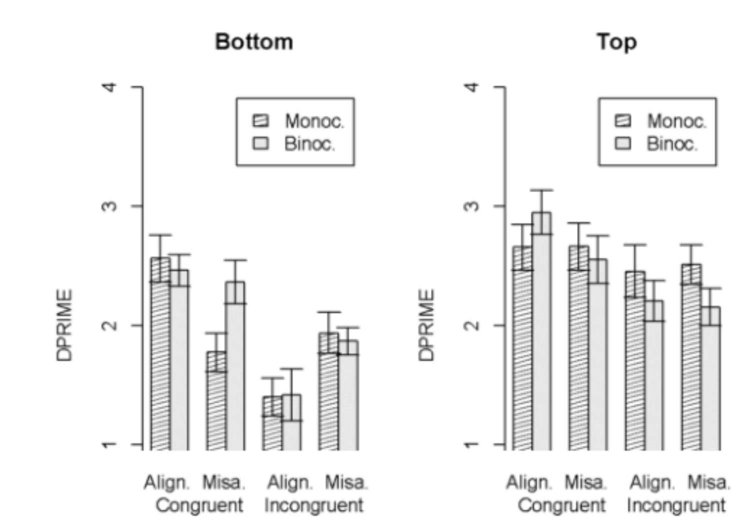


Figure 5a This barplot shows the four-way (highest-order) interaction, Eye x Alignment x Congruency x Side, which approaches significance: F=3.33,1, df=(1, 299), p=0.0691. Barplots of average d' values by condition combinations, separated by location of cue (Bottom/Top). Error bars denote one standard error. Monocular presentation has larger differences in d' between the congruent/incongruent conditions than binocular, demonstrating monocular facilitation.

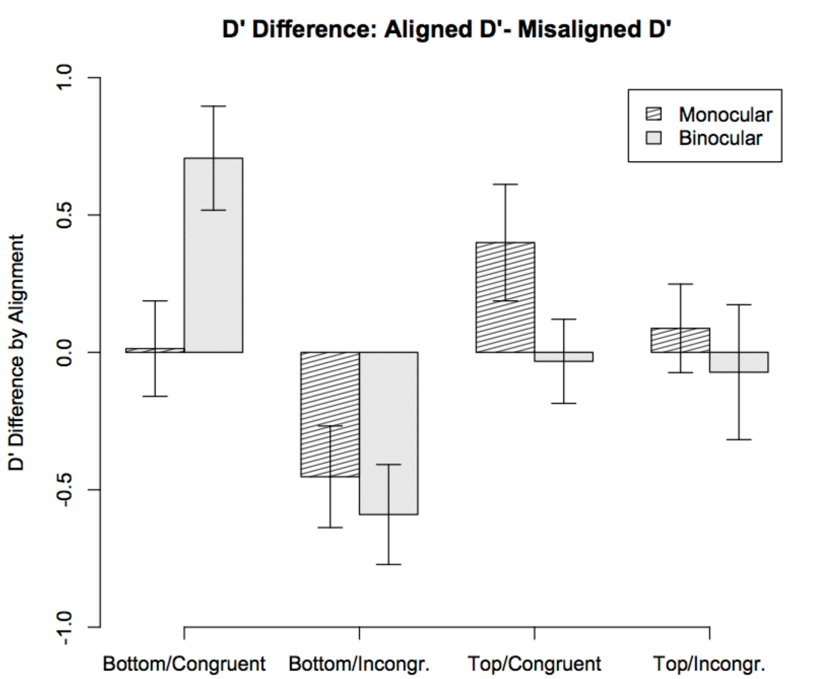


Figure 5b There is a decrease in d' in aligned over misaligned conditions when the faces are incongruent—that is, there is an increase in false alarms for both monocular and binocular presentation, only when the bottom is cued. This suggests that participants are more influenced by the top of the face (that is, the top is more likely to interfere with focus on the cued bottom than vice versa).

To investigate this four-way interaction further, we performed paired t-tests (Table 2). Since this was post-hoc testing, a Bonferroni adjustment was performed for significant p values. After this adjustment for multiple testing, we lost significance of the monocular, cued bottom, incongruent combination (Mean difference = -.453, t = -2.448, df = 19, p = .024, adjusted p = .194) determining that aligned trials under such conditions had noticeably lower discrimination sensitivity, which would have been the typical composite face effect though only when the bottom of the face was cued. However, binocular and bottom-cued trials with both congruencies remained showed significant differences based on alignment: for congruent trials, aligned trials showed a significant improvement over misaligned (Mean difference = .707, t = 3.733, df = 19, p = .001, adjusted p = .011) while this effect was the opposite for incongruent trials: as the composite face effect would predict, interference was highest on the aligned trials (Mean difference = -.590, t = -3.245, df = 19, p = .004, adjusted p = .034). Taking the 4-way interaction from Table 1 as near significant, and noting the two-way interaction later discussed in figure 8, we posit monocular facilitation of the CFE.

Eye	Cued Side	Congruency	Mean diff	t	df	p	Adjusted p
Monocular	Top	Congruent	0.400	1.885	19	0.075	0.599
Monocular	Top	Incongruent	0.087	0.543	19	0.594	1.000
Monocular	Bottom	Congruent	0.014	0.079	19	0.938	1.000
Monocular	Bottom	Incongruent	-0.453	-2.448	19	0.024	0.194
Binocular	Top	Congruent	-0.032	-0.212	19	0.834	1.000
Binocular	Top	Incongruent	-0.072	-0.294	19	0.772	1.000
Binocular	Bottom	Congruent	0.707	3.733	19	0.001	**0.011**
Binocular	Bottom	Incongruent	-0.590	-3.2	19	0.004	**0.034**

Table 2: Paired Sample t-test for difference in D' Based on alignment: monocular presentation shows a CFE trend (aligned impaired more than misaligned for incongruent trials), as does binocular, but binocular presentation displayed an additional advantage for congruent trials with aligned trials.

The two highest-order significant interactions were a 3-way interaction between the cued side, alignment and congruent condition (F=6.32, df=(1, 299), p=0.0125) and the two-way interaction between monocular/binocular presentation and the congruent/incongruent condition (F=5.36, df=(1, 299), p=0.0213).

An inspection of the model coefficients revealed that, adjusted for all other factors in the model, there were no significant differences in d' values between monocular/binocular condition for the congruent condition (b=0.17, SE=0.10, t=1.65, df=305, p=0.1004), while in binocular presentation, the d' values for incongruent trials were significantly lower than those for congruent trials; however, the interaction term indicated that for the incongruent condition, monocular presentation resulted in lower than expected d' value (interaction term b=-0.34, SE=0.15, t=-2.30, df=305, p=0.0221).

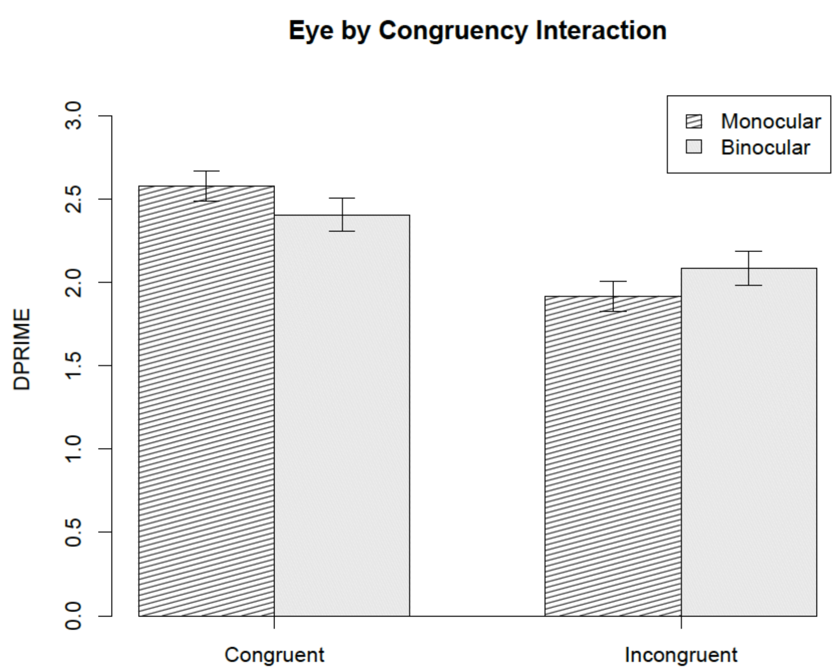


Figure 8: Significant two-way interaction between Eye and Congruency: monocular presentation resulted in lower than expected d' value on incongruent trials, while binocular did not. (interaction term b=-0.34, SE=0.15, t=-2.30, df=305, p=0.0221). In other words, the effect of congruency is greater for monocular presentation.

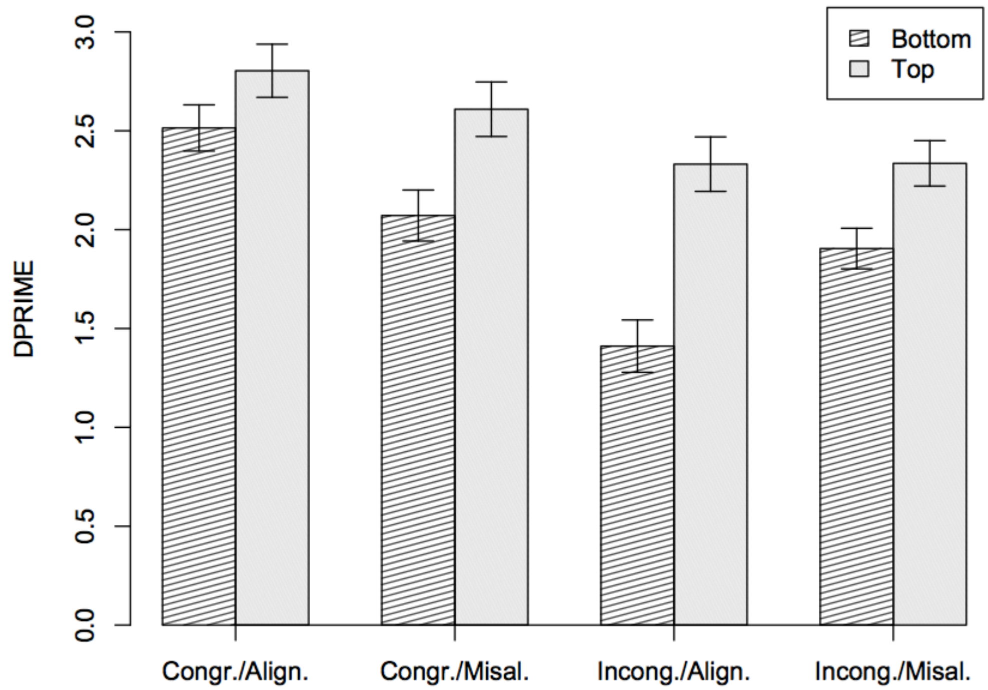


Figure 9: Significant three-way interaction of Cued Half x Alignment x Congruency: Aligned face presentation leads to increased sensitivity in the congruent condition, and decreased in the incongruent condition, a typical CFE effect. The effect is more extreme when the bottom is cued, which shows an inability to filter out the information from the incongruent top half, implying an HP effect from the top of the face but not the bottom.

Specific visual field presentation (left/right) was not taken into account in the final model because with only considering the main effect model , the 4-level LL-LR-RL-RR variable only has a trend-level effect on d' (F = 2.33, df = (3, 644), p = 0.0730), and when we looked at the contrasts between the levels, using Tukey's post-hoc adjustment, none of the differences were even on a trend level (Table 3).

	Estimate	Std. Error	z value	Pr(>|z|)
LR - LL == 0	0.024900	0.101939	0.244	0.995
RL - LL == 0	-0.178667	0.101939	-1.753	0.296
RR - LL == 0	-0.176114	0.101939	-1.728	0.309
RL - LR == 0	-0.203568	0.101939	-1.997	0.189
RR - LR == 0	-0.201014	0.101939	-1.972	0.199
RR - RL == 0	0.002553	0.101939	0.025	1.000

Table 3: Tukey Contrast shows no differences in d' for lateralization: neither visual stream shows a significant advantage over the other in face recognition in this study. "LL" refers to a trial where both the study and test face are presented to the left eye, while "RL" would represent a trial where the study face was on the right and the target face was on the left.

Discussion

Combining paradigms used to behaviorally test subcortical processing and holistic face processing, this study aimed to examine a potential subcortical contribution to HP by checking for increase in false alarms compared to correct responses while identifying composite faces. Through use of a mirror stereoscope to present separate images to each eye, thereby isolating monocular pathways, and the use of the composite face task to measure interference due to HP, we aimed to show that the “false alarm” rate on incongruent trials would be greater with monocular presentation than binocular presentation due to the “pull” toward incorrect answers by HP. The evolutionary implication is that HP for faces is not specific to advanced neocortical structures and may in fact be an older mechanism than the neocortex itself.

Indeed, the effects of congruency were more pronounced in monocular presentation, showing a greater failure in selective attention. However, the full composite effect—aligned, incongruent trials leading to decreased performance—was only found when the bottom was cued, and therefore the HP interference was coming from the top of the face. There was also a strong facilitation trend observed for monocular presentation, although the analysis lacked sufficient power to report significance. Overall, the composite face effect was replicated in both monocular and binocular conditions when the bottom was cued, and displayed a slight monocular advantage.

In light of these trends, this experiment should be repeated with other measures of holistic processing, and perhaps applying the composite task to inanimate objects to compare to faces. More in-depth analysis of reaction time or inverse efficiency may also be of interest. Additionally, testing prosopagnosic individuals, who generally cannot display HP of faces, would provide insight into potential subcortical compensation or contribution after cortical damage to typically face-coded areas.

Further research into subcortical lateralization is also warranted; although we did not find a significant effect for left or right eye presentation, the lateralization of certain functions like word or face processing in the cortex might also affect the results even during segregation of the two visual fields: including data from left-handed individuals might make this even more interesting.

More work should also be done on distinguishing between the effects of the top and bottom of the face--it seems that the halves contribute disproportionately for holistic processing. It is possible that we simply receive more information from the top and therefore have more difficulty ignoring it: a region-specific account of holistic processing should be explored. Wang et al., while presenting the composite task to an Asian population using Asian, Caucasian, and monkey faces, found a significant same-species and same-race advantage for the composite face effect for processing the top of the face, but when focusing on the lower half, the composite face effect when studying monkey faces was as strong as for both Asian and Caucasian faces. The implication is that inter-species holistic processing of faces is limited to the top of the face, while both halves of the face are holistically processed within species (Wang et al. 2018). Children with autism spectrum disorder have been shown to perform similarly to controls when making discrimination judgments based on the mouth, but show a significant deficit in the same task focusing on the eyes (Wolf et al. 2008). While individuals with autism do not show significant deficits in HP for faces, they do show an advantage for local over global processing tasks (Tanaka et al. 2016). Other manipulations might involve spatial frequency.

Ultimately, these results call for further evolutionary neuropsychology research into subcortical structures' influence on holistic processing, as some aspect of HP is likely done before advanced cortical layers previously implicated. While we did not expect to find such intense effect differences between the top and the bottom of the face, it also lends insight into the fact that HP might require access to the top of the face and be region-specific. Breaking down the way we identify faces, constantly, every day, not only furthers evolutionary psychology and perception literature, but also might help create better face-recognition software and deepen our understanding of human deficits in face processing.

References

Boremanse A., Norcia A. M., Rossion B. (2013). An objective signature for visual binding of face parts in the human brain. Journal of Vision, 13 (11): 6, 1–18.

Cauchoix, M., & Crouzet, S. (2013). How plausible is a subcortical account of rapid visual recognition? .*Frontiers in Human Neuroscience* . Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2013.00039

Gabay, S., Nestor, A., Dundas, E., & Behrmann, M. (2014). Monocular Advantage for Face Perception Implicates Subcortical Mechanisms in Adult Humans. *Journal of Cognitive Neuroscience*, *26*(5), 927–937.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the Middle Fusiform "Face Area" Increases With Expertise in Recognizing Novel Objects. *Nature Neuroscience*, *2*, 568. Retrieved from http://dx.doi.org/10.1038/9224

Isabel Gauthier, Marlene Behrmann, and Michael J. Tarr (1999). Can Face Recognition Really be Dissociated from Object Recognition? Journal of Cognitive Neuroscience, Vol. 11, No. 4 , Pages 349-370 (doi: 10.1162/089892999563472)

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect?. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 628.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.

Lin, I.-R., & Chiao, C.-C. (2017). Visual Equivalence and Amodal Completion in Cuttlefish. *Frontiers in Physiology*, *8*, 40. http://doi.org/10.3389/fphys.2017.00040

Liu, T. T., & Behrmann, M. (2014). Impaired holistic processing of left-right composite faces in congenital prosopagnosia . *Frontiers in Human Neuroscience* . Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2014.00750

Menon RS, Ogawa S, Strupp JP, Uğurbil K. Ocular dominance in human V1 demonstrated by functional magnetic resonance imaging. Journal of Neurophysiology. 1997;77:2780–2787.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1-19.

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

Tamietto, Marco, et al. "Collicular vision guides nonconscious behavior." *Journal of Cognitive Neuroscience* 22.5 (2010): 888-902.

Tanaka, J. W., & Sung, A. (2016). The "Eye Avoidance" Hypothesis of Autism Face Processing. *Journal of Autism and Developmental Disorders*, *46*(5), 1538–1552. https://doi.org/10.1007/s10803-013-1976-7

Wang, Z., Quinn, P. C., Jin, H., Sun, Y.-H. P., Tanaka, J. W., Pascalis, O., & Lee, K. (2018). A Regional Composite-Face Effect for Species-Specific Recognition: Upper and Lower Halves Play Different Roles in Holistic Processing of Monkey Faces. *Vision Research*. https://doi.org/10.1016/j.visres.2018.03.004

Wolf, J. M., Tanaka, J. W., Klaiman, C., Cockburn, J., Herlihy, L., Brown, C., … Schultz, R. T. (2008). Specific impairment of face processing abilities in children with autism spectrum disorder using the Let's Face It! Skills Battery. *Autism Research : Official Journal of the International Society for Autism Research*, *1*(6), 329–340. https://doi.org/10.1002/aur.56