"I Can't See Your Eyes Well 'Cause Your Nose is Too Short": An Interactivity Account of Face Processing

Reza Farivar (reza.farivar@mail.mcgill.ca)
Dept. of Psychology, McGill University, 1205 Dr. Penfield Ave.
Montreal, QC, H3A 1B1, Canada

Avi Chaudhuri (avi@hebb.psych.mcgill.ca)
Dept. of Psychology, McGill University, 1205 Dr. Penfield Ave.
Montreal, QC, H3A 1B1, Canada

Abstract

The present work utilizes the generalized form of the signal detection theory (the General Recognition Theory) to formally model representation of faces during face perception. We tested the hypothesis that face perception, typically described as a holistic or configural process, can be formally described as an interactive processing of face parts, whereby one component of a face influences perception of other components. We present theoretical and experimental developments on this topic, building on previous work, but utilizing more realistic stimuli, a powerful mathematical model, and a crucial comparison condition.

Introduction

Why can a touch of makeup can make a nose smaller, lips chubbier, and eyes larger? Why does a goatee make a face longer, or a moustache more round? Face components must be processed interdependently, one influencing how the other is perceived.

How can this process be best described?

Two general frameworks for answering this question have been proposed. One posits that faces are processed largely in terms of the geometric relationship of their features. Local features such as eyes, nose, and mouth, are distinguished from geometric features or spatial distributions of features; both are thought to contribute to the processing of faces (Leder & Bruce, 2000; Searcy & Bartlett, 1996; Diamond & Carey, 1986). Another recent approach is to define face processing as a holistic-based process; the representation of faces is thought to be less part-decomposed than other objects (Farah, et al., 1998; Tanaka & Farah, 1993). The predictive and explanatory difference between these approaches is currently in debate.

A related approach has been to formalize face processing as an interactive process (Sergent 1984; Macho & Leder 1998; Thomas 2001). Such formalization depicts face-specific effects as resulting from interdependencies in the processing of various components of a face. For example, it may be conceived that the representation of a face cannot be varied on only one dimension, but must be varied on multiple dimensions at once, because the various dimensions are dependent. This is in contrast to a non-interactive system, where a representation can vary on a given dimension independently of its variance on other dimensions; an analogy can be made in relation to the definition of an interaction in relation to the General Linear model, where the interaction term is a conditional (or dependent) function of at least two dependent variables.

Sergent (1984) presented subjects with Identikit stimuli that varied on a number of dimensions including facial contour, eyes, and internal spacing. By regressing reaction times (RTs) on same-different judgments, she observed an interactive influence between the components of internal spacing and contour, and through multidimensional scaling, she noted that dissimilarity judgments deviated from a perfect cube (which would be expected if the judgments were based on interaction of the stimuli dimensions). Moreover, these effects were observed for upright and not inverted faces. Although her formal approach was quite powerful, factors relating to (a) the regression analysis, (b) realism of the stimuli, and (c) the lack of a statistic for the multidimensional scaling data (see Macho and Leder 1998), make her results difficult to interpret.

Macho and Leder (1998) tested the interactivity hypothesis more carefully using realistic stimuli, and employing the logit model to model their data. Using this approach, the researchers did not observe an interactivity effect. This lack of an effect may in part be due to the choice of model they used for their data. The logit model does not differentiate between many interactions that are possible. Also, the logit model makes the assumption that decisional effects in perceptual tasks are non-existent, and this assumption is not validated. For example, the frequent violation of this assumption is at the heart of developments in signal detection theory and its generalization, the General Recognition Theory (Maddox, 1992). Another possible explanation of those results may be that the response set of the participants was impoverished, since they were matching 27 possibilities to 2 targets.

The General Recognition Theory.

Signal Detection Theory (SDT; Swets, 1996) has been a very successful and widely utilized model for separating decisional and perceptual effects in detection, identification, or categorization tasks. The theory assumes that the perceptual representation of a given stimulus varies from trial to trial, and this variability can be represented by a Gaussian distribution. Noise in the system is also considered to follow such a pattern, resulting in two overlapping Gaussian distributions. In order to make a judgment, one has to place a decisional boundary; stimuli perceived as being above this boundary are called signals, and those below are called noise. Using probabilities of hits and false-alarms, the distance of these distributions (called d') can be estimated, as can the criterion (C). Intuitively, the distance between the two distributions is a good measure of perceptual sensitivity to signal and noise—the greater the distance between the two distributions, the less likely a confusion will be made between signal and noise.

However, the SDT is unidimensional, meaning that it can be only applied to analysis of stimuli varying on a single dimension. The majority of stimuli in our environment are obviously multidimensional and for this reason, a generalization of this theory to multiple dimensions would be of great use; the GRT is the generalized form of SDT. Consider a two-dimensional system: faces varying on eye-distance and nose-length, and two levels on each of the two dimensions. The GRT would represent the perception of each of the four stimuli as two-dimensional Gaussian distributions. The segregation or distance of each of the four distributions from the other would represent the discriminability of those stimuli, and decisional boundaries can be drawn between these distributions. However, two-dimensional distributions can also vary in shape or orientation, and this has implications for the underlying perceptual representations.

If, within a stimulus, the perception of one dimension is dependent on perception of the other dimension, Perceptual Dependence is said to exist for those dimensions within that stimulus. From a topographic view of a two-dimensional Gaussian distribution, this dependence is represented as a tilted ellipse; it is analogous to a correlation and reflects that identification errors for that stimulus are not separable into errors on a given dimension, but are errors on two dimensions simultaneously.

The spatial arrangement of these Gaussian distributions in the perceptual space can also be varied. Variations of this sort reflect variations in d's between the Gaussian distribution. In the terminology of the GRT, when d' for one dimension varies as a function of the levels of the other dimension, the dimension in question is said to be Perceptually Inseparable from the other dimension. Note that the converse is not necessarily true.

Finally, the spatial arrangement of the decisional boundaries in this perceptual space may also vary; when the position of the decisional boundary for one dimension varies across levels of the other dimension, the dimension in question is said to be Decisionally Inseparable from the other dimension. Again, the converse is not necessarily true.

Within the GRT framework, the notion of holistic face processing may be formalized as either perceptual dependence or inseparability. This is because if either perceptual separability or independence fail, then it suggests that the perception of one aspect of the stimuli influences the perception of the other aspect of the stimuli, which is more simply called an interaction of stimulus dimensions.

The strengths of the GRT led Thomas (2001) to use this powerful model to test the interactivity hypothesis, but due to aspects of the methodology, interpretation of the results is limited. First, the stimuli used in that study were artificial—they were schematic line drawings of a face, lacking texture or asymmetry. Recent evidence suggests that line drawn images of faces are more difficult to remember and do not give rise to a configural code as photographic images of faces do (Leder, 1999; Leder, 1996). The possibility may thus exist that stimuli used in Thomas (2001) do not tap into face processing the same way that face photographs do.

Second, she used four subjects—two for each pair of dimensional manipulations. Although it is common to use few subjects in studies of perception, it may be the case that a high-level visual process, such as face perception, is more variable between subjects. Thus it may be the case that the subjects tested did not display the phenomenon for the stimuli used. Finally the control used in Thomas (2001) does not directly allow us to compare normal face processing with a baseline, such as the processing of inverted faces.

Taken together, a reliance on past studies that tested the interactivity hypothesis is hindered by problems such as (a) small sample size, (b) poor stimuli, (c) insensitive or inappropriate statistical modeling, or (d) lack of a proper control condition.

In this study, we asked (a) whether face processing can be represented as interactive processing of face components, (b) whether different interactions are dissociable, and (c) whether interactive processing uniquely occurs for upright faces. We present results suggesting that face processing can indeed be represented as interactive processing of face components, that a number of dissociable interactions can be observed, and that the observed interactions are unique for the processing of upright faces

Methods

Participants

Forty-seven Psychology graduate and undergraduate students participated in this study. Each either received bonus course credit or was financially compensated. All had normal or corrected-to-normal vision. Twenty-four participants viewed upright faces, while twenty-three viewed inverted faces.

Stimuli

A stimulus set was constructed according to a feature-complete factorial combination of eye-to-eye distance (short vs. long distance) and nose length (short vs. long nose). A single grey-scale photograph of a male face served as the base stimulus, and the sets were derived from manipulations of this face. All manipulations were made digitally using Adobe Photoshop 5.0. This medium of manipulation ensured that the faces were identical in all other aspects (contrast, brightness, texture, etc.) except for the manipulated features.

For the inverted-face condition, the same set of faces were used, but inverted. The stimuli are illustrated in Figure 2.

Figure 2. Example of stimuli used in this study

Apparatus

The experiment was conducted with a Macintosh computer, using the Psychtoolbox for Matlab (Brainard 1997; Pelli 1997). Participants were seated approximately 40 cm away from the screen, giving the images a visual angle of approximately 4 degrees in width and 6 degrees in height. Responses were collected on a computer keyboard, using the numeric keypad or the row of number keys on the main keypad.

The experiment took place in a quiet and dimly lit environment.

Procedure

Participants were informed that they would view four face images, and that the face images would differ only slightly from one another, and as such they should pay careful attention to the small differences to properly complete the task.

Each experimental block consisted of 100 trials, with each version of the face being presented 25 times. The presentation was randomized, with the restriction that the same version of the face would not be viewed more than two times consecutively. Each trial began with the presentation of a '+' cue, which appeared at a location equal to the center of the target face. The cue was present for 200 msecs, and was followed by a 200 msec delay, after which the target face appeared for 125 msecs. We used this short presentation time because we did not want subjects to have the opportunity of analyzing each component separately. A half-second delay followed the target face, and subsequently the four possible test faces appeared.

The test faces were all the possible versions of the face that appeared for that experiment. The location of the test faces was randomized on each trial. The participants had to make an identification judgment for the target face by selecting one of the four possible responses (i.e., matching/identification task). The testing phase was not timed, but participants were encouraged to make their response within five seconds.

A session consisted of four blocks and lasted between 45 minutes to 1 hour. After each block, participants were given feedback on their performance for that block and were then given the occasion to take a break. Also, during the experiment, after every 20 presentations, a brief break was offered by an on-screen prompt.

Data and Results

Each subject's responses were collected in a 4x4 confusion matrix. The matrices for each condition were collapsed across all subjects in the condition before subjecting them to analyses.

Table 1 – d' and C estimates for the dimension of nose length across eye distance for upright and inverted faces (~ denotes negation)

Nose Length Across Eye Distance				
	Upright		Inverted	
Eye Distance	$d'_{Nose\ Length}$	$C_{Nose\ Length}$	$d'_{Nose\ Length}$	$C_{Nose\ Length}$
Eyes Close	1.849	0.886	1.100	0.543
Eyes Far	1.680	0.757	1.147	0.538
$Z_{observed}$	2.858**	3.139**	0.842	0.129
Conclusions	~PS	~DS	PS	DS

Table 2 – d' and C estimates for the dimension of eye distance across nose length for upright and inverted faces. (~ denotes negation)

Eye Distance Across Nose Length				
	Upright		Inverted	
Nose Length	$d'_{Eye\ Distance}$	$C_{Eye\ Distance}$	$d'_{Eye\ Distance}$	$C_{Eye\ Distance}$
Short Nose	1.663	0.801	1.666	0.788
Long Nose	1.807	0.843	1.623	0.729
$Z_{observed}$	2.439**	1.028	0.733	1.450
Conclusions	~PS	DS	PS	DS

Using MSDA-2 (Kadlec 1995), the data were subjected to multidimensional signal detection analyses to make estimates of the different types of interactions– Perceptual Separability, Perceptual Independence, and Decisional Separability. All tests were two-tailed Z-tests, with $\alpha = 0.05$. Tables 1 and 2 present the d' and C estimates in the macroanalyses pertaining to Perceptual and Decisional Separabilities. It should be noted that the multidimensional signal detection analysis approach makes the assumption of normality of the perceptual distributions and variance equality amongst those distributions.

The reader is referred to Kadlec (1995) and Kadlec and Townsend (1992; 1992) for full details of the analysis. Briefly, the analytic method involves a macroanalysis and a microanalysis, which together reveal information about perceptual and decisional separability. In the macroanalysis, traditional SDT estimates of d' (a measure of sensitivity) and C (an estimate of a decisional boundary, which tells us about decisional biases) are made on one dimension across one level of another dimension. This results in d' and C estimates for each dimension at every level of the other dimension. The values are compared using a Z test—significant difference between d' or C of one dimension across levels of another dimension suggests a violation of Perceptual and Decisional Separability, respectively, and the direction of this interaction can readily be ascertained by looking at the d' and C estimates.

The tilt of the individual perceptual distributions can be estimated by the information provided in the microanalysis, where conditional d' values are measured for each stimulus (Kadlec, 1995). For example, if the d' for eye distance is larger for short-nose stimuli that are properly identified as short-nose than for short-nose stimuli judged as long-nosed, then one or both of the short-nosed stimuli have tilted perceptual distributions—i.e., perceptual independence fails in one or both. By estimating all possible conditional d' in this way, the tilts of the distributions can be estimated. This analysis further elucidates results of the test of Sampling Independence (Ashby & Townsend, 1986), which identifies cases where Perceptual Independence may have failed.

Tables 1 and 2 show that for upright faces, perceptual separability failed for both dimensions of eye-to-eye distance and nose length, but this pattern was not observed for inverted faces. This suggests that the ability to discriminate between levels of each dimension was dependent on the levels of the other dimensions—that the dimensions interact to bias perception.

For upright faces, discrimination of eye distance was significantly better when the nose was longer, while discrimination of nose length was significantly better when the eyes were close to each other. For inverted faces, however, discrimination ability for one dimension was not influenced by changes in the other dimension. Furthermore, for upright faces, a significant bias was observed in the macroanalysis in judging nose length across levels of eye distance—for eyes close to one another, participants were biased towards judging the face as having a longer nose, while for eyes far apart, they were biased to judge a face as having a shorter nose. No strong support for decisional separability was obtained in the microanalysis for upright or inverted faces, suggesting decisional interactions to take place in both conditions.

The microanalysis, combined with the tests of sampling independence suggested that perceptual independence failed on several occasions in upright faces, but failed less so for inverted faces; it should be noted that based on the GRT constructs, given that decisional separability failed for upright faces, it is difficult to draw any firm conclusions relating to perceptual independence (Kadlec and Townsend 1992). However, assuming the failure of perceptual independence to be true, this type of interaction can best be thought of as an association or a Gestalt effect. These results suggest that for upright faces, a face with eyes close together is perceived as having a longer nose, while a face with eyes far apart is perceived as having a shorter nose. These results show that different interactive processes may be uniquely recruited for the processing of upright faces.

Discussion

General Discussion

Our results show that face processing can indeed be modeled as an interactive process, being especially valid for upright faces. The results corroborate previous findings (Leder, 1996; Tanaka & Sengco, 1997; Farah, et al., 1998), and build on previous modeling attempts (Sergent, 1984; Macho & Leder, 1998; Thomas, 2001).

Most recently, Wenger & Ingvalson (2003) investigated face perception in a manner quite similar to our approach here, but our results are different. Although the authors also made use of realistic stimuli, they did not observe perceptual interactions between facial dimensions for upright or inverted faces; only a decisional component was observed. Our experiment, however, differs with their study in at least one respect: that individual stimuli were presented for very brief periods of time (125 ms) in our study, but for 3 s in Wenger and Ingvalson (2003). The possibility may thus exist that subjects had time to investigate individual features more independently and this may have resulted in fewer perceptual interactions of face components

Model Sensitivity and Methodological Considerations

The results presented here point to important considerations that need to be made with respect to choice of model and experimental design. Firstly, it is quite likely that Macho and Leder (1998) were unable to observe interactive influences due to their choice of analytic model. The experiment presented here is quite similar to theirs in terms of stimuli used and manipulations made, but the results are different, showing important patterns of interactions to uniquely underlie face processing. Secondly, although Thomas (2001) did use the same model the lack of detail in the stimuli may have undermined the possibility of observing an interaction if it was present (Leder, 1996; 1999). Our study is quite analogous to hers, but we have collapsed the data across a large number of subjects and have used photographic images of faces for ours stimulus set. It is likely that the use of photographic images induces a more configural/holistic type of processing, thereby giving rise to the observed interactions (Leder, 1996; 1999). The fact that we have tested a large number of subjects may have also increased our sensitivity to the effect by reducing the variability that may be present between subjects.

A potential criticism of the current work relates to the use of the inverted-face condition as a control. A number of different control conditions may be used, such as the one used in Thomas (2001) where the presented images do not vary on any dimensions. Another possibility is to make use of another class of objects, such as animal faces, as control. However, using images that are unchanged does not allow for integration into the GRT, and thus a direct comparison cannot be made between face images that were varied on the given dimensions versus those that were not. Use of a different class of objects gives rise to the added difficulty that the difference between the experimental and control conditions may no longer reflect a difference between normal face processing and non-face processing, rather the difference may be due to a number of extraneous differences between face and non-face images. The comparison between inverted and upright faces is made here because the inverted face images contain exactly the same information as upright face images but are not processed in the same manner as upright faces (Leehey, et al., 1978; Leder, et al., 2001).

Types of Interactions

We have found that perception of one part of the face does influence perception of other parts of the face. We have reported two interactions with such a relationship to be mostly unique to the viewing of upright faces. What are the differences between these interactions, and what do they tell us about face processing?

Perceptual dependence, as defined and used here, is primarily a within-stimulus effect (i.e., within a single stimulus, perception of one component may interact with the perception of the other component). Such an effect has been previously ascribed to emergent properties, or a Gestalt dimension (Ashby & Townsend, 1986; Kadlec & Townsend, 1992) and is perhaps a "strong" representation of a holistic process (Farah, et al., 1998). If Perceptual Dependence demonstrates Gestalt or holistic processing, then how is one to interpret Perceptual Inseparability? Is it a different effect than that of holistic processing? Is this analogous to configural processing? Some researchers would argue that such an effect points to configural processing (Macho & Leder, 1998; Leder & Bruce, 2000; Leder, et al., 2001) while others may believe that it corresponds to holistic processing (Tanaka & Farah, 1991; Farah, et al., 1998). How exactly such an effect is to be interpreted is dependent on the view one takes. On the one side, such an effect can be considered a configural effect because it relates to how perception of one component (i.e., nose length) is influenced by variation in the geometric position of another component (i.e., eye distance) and vice versa. However, the holistic hypothesis can be reconciled here as well (i.e., Tanaka & Sengco, 1997) with the results—e.g., in a holistic process, all parts and components are more integrated, thus resulting in a perceptual inseparability effect.

Perhaps the results can be interpreted and extended using a different paradigm all together. Our results suggest that there may be two separable effects here: a within-stimulus component (perceptual dependence) and a between-stimulus component (perceptual inseparability). Both appear to be largely related to the appropriate processing of upright faces, and the two effects are independent of one another. It is therefore possible that face processing, whether holistic or configural, involves to some extent both types of interactions—certain tasks may tap more into "between stimulus" component (i.e., Tanaka & Sengco, 1997; Farah, et al., 1998) whereas other tasks may tap into the "within-stimulus" component (i.e., Tanaka & Farah, 1993). It is quite likely that different manipulations may be used to better understand these two parallel processes and a number of research questions can be asked in relation to this finding: does attention modulate the within- or the between-stimulus interactions, or both? Do these interactions change over time towards greater integration? Do we acquire such an integration for all visual objects for which we have developed expertise? What is the relationship between physical dimensions of faces as estimated by techniques such as principal components analysis (Valentine, 1991) and perceptual interactions?

Conclusions

On the basis of our findings, we suggest that faces are indeed represented and/or processed in an interactive manner that is compatible with a "strong" version of the holistic hypothesis (Farah et al., 1998). However, some aspects of our findings can also be explained using the configural process, suggesting that these paradigms alone may be limited in explaining face processing in the brain. Our results, using a powerful formal model suggest a different pair of parallel processes to be involved at least in the on-line processing of faces. We have shown these effects to be largely unique to upright faces and not inverted ones, a finding that is of importance because it suggests an interactive mode of processing of complex visual stimuli representing an upright face.

Acknowledgements

This research was supported by an NSERC PGS-B fellowship awarded to RF and an NSERC operating grant awarded to AC. The authors wish to thank Isabelle Boutet and Chang Hong Liu for their helpful comments on various stages of the study.

References

Ashby, F. G. and J. T. Townsend (1986). "Varieties of perceptual independence." *Psychological Review 93*, 154-79.

Brainard, D. H. (1997). "The Psychophysics Toolbox." *Spatial Vision 10*, 433-6.

Diamond, R. and S. Carey (1986). "Why faces are and are not special: an effect of expertise." *Journal of Experimental Psychology: General, 115*, 107-17.
Farah, M. J., K. D. Wilson, et al. (1998). "What is "special" about face perception?" *Psychological Review, 105*, 482-98.
Kadlec, H. (1995). "Multidimensional signal detection analyses (MSDA) for testing separability and independence: A Pascal program." *Behavior Research Methods, Instruments, & Computers ,27,* 442-458.
Kadlec, H. and J. T. Townsend (1992). "Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions." *Journal of Mathematical Psychology 36*, 325-374.
Kadlec, H. and J. T. Townsend (1992). Signal detection analyses of dimensional interactions. *Multidimensional models of perception and cognition. Scientific psychology series*. F. G. Ashby. Hillsdale, NJ, Lawrence Erlbaum Associates, Inc**:** 181-227.
Leder, H. (1996). "Line drawings of faces reduce configural processing." *Perception 25*, 355-66.
Leder, H. (1999). "Matching person identity from facial line drawings." *Perception, 28*, 1171-5.
Leder, H. and V. Bruce (2000). "When inverted faces are recognized: the role of configural information in face recognition." *Quarterly Journal of Experimental Psychology: A, 53,* 513-36.
Leder, H., G. Candrian, et al. (2001). "Configural features in the context of upright and inverted faces." *Perception, 30,* 73-83.
Leehey, S., S. Carey, et al. (1978). "Upright and inverted faces: the right hemisphere knows the difference." *Cortex, 14*, 411-9.
Macho, S. and H. Leder (1998). "Your eyes only? A test of interactive influence in the processing of facial features." *Journal of Experimental Psychology Human Perception and Performance, 24,* 1486-500.
Maddox, W. T. (1992). Perceptual and Decisional Separability. *Multidimensional Models of Perception and Cognition.* F. G. Ashby. Hillsdale, New Jersey, Lawrence Erlbaum Associates**:** 147-180.
Pelli, D. G. (1997). "The VideoToolbox software for visual psychophysics: transforming numbers into movies." *Spatial Vision, 10*, 437-42.
Searcy, J. H. and J. C. Bartlett (1996). "Inversion and processing of component and spatial-relational information in faces." *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904-15.
Sergent, J. (1984). "Configural processing of faces in the left and the right cerebral hemispheres." *Journal of Experimental Psychology: Human Perception and Performance, 10,* 554-72.
Sergent, J. (1984). "An investigation into component and configural processes underlying face perception." *British Journal of Psychology, 75,* 221-42.
Swets, J. A. (1996). *Signal Detection Theory and Roc Analysis in Psychology and Diagnostics: Collected Papers.* Mahwah, N.J., Lawrence Erlbaum Assoc.
Tanaka, J. W. and M. J. Farah (1991). "Second-order relational properties and the inversion effect: testing a theory of face perception." *Perception and Psychophysics, 50,* 367-72.
Tanaka, J. W. and M. J. Farah (1993). "Parts and wholes in face recognition." *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 46A,* 225-245.
Tanaka, J. W. and J. A. Sengco (1997). "Features and their configuration in face recognition." *Memory and Cognition, 25,* 583-92.
Thomas, R. D. (2001). "Perceptual interactions of facial dimensions in speeded classification and identification." *Perception and Psychophysics, 63,* 625-650.
Valentine, T. (1991). "A unified account of the effects of distinctiveness, inversion, and race in face recognition." *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology 43A,* 161-204.
Wenger, M.J., and Ingvalson, E.M. (2003). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 872-892