"I can't see your eyes well 'cause your nose is too short": An interactivity account of holistic and configural face processing

Reza Farivar

Department of Psychology

McGill University

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Masters of Science

May, 2002

Your file Votre référence
ISBN: 0-612-85784-0
Our file Notre référence
ISBN: 0-612-85784-0

Acknowledgements

Special thanks to my supervisor, Avi Chaudhuri, for his support throughout this project. Also thanks goes to a number of my lab colleagues, especially Alain Mignault, Chang Liu, Karen Borrmann, Cary Kogan, and Isabelle Boutet for the suggestions on various aspects of the project. Thanks are due also to Marie St. Laurent for her help in collecting some of the data, and to Carmelo Milo for general support throughout the study.

Table of Contents

List of Tables and Figures

Abstract

What is meant by configural and holistic processing? The present project attempts to answer this question by formalizing configural and holistic processing as interactive processing of face parts. Four face images were created from one main face by varying the eye distance and nose length of the face to yield a 2x2 feature-complete factorial combination set of stimuli. Participants viewed each version of the face for 100 msecs, and then identified the face they saw. Their responses were subjected to multidimensional signal detection analysis to obtain estimates of different types of perceptual interactions defined by General Recognition Theory (Ashby & Townsend, 1986). It is shown here that perception of upright faces exhibits a number of interactions that are not present for inverted faces. The nature of these interactions are linked to concepts of holistic and configural face processing. A computational justification is forwarded for this interactivity account of face processing.

Résumé

Quatre différents visages furent générés à partir d'un visage de départ en manipulant la distance inter-occulaire et la longueur du nez. Une combinaison de stimuli factorielle (2X2) à caractéristiques complètes fut produite. Lors de la tâche expérimentale, les participants visionnaient l'un des quatre visages modifiés pendant 100ms et identifiaient celui qui leur était présenté. Leurs réponses furent soumises à une analyse multidimentionnelle de détection de signal; une estimation des différents types d'interactions perceptuelles définies par la Théorie Générale de Reconnaissance (Ashby et Townsend, 1986) fut obtenue. Il fut démontré que la perception de visages non-inversés entraîne certaines interactions perceptuelles qui sont absentes lors de la perception de visages inversés. La nature de ces interactions est compatible avec les concepts de traitement holistic et de traitement configurationel des visages. Une justification computationelle est avancée afin de décrire le traitement interactionnel des visages.

Introduction

It has recently been proposed that faces are processed holistically, such that each face is represented as an undecomposed whole (Farah, Wilson, Drain, & Tanaka, 1998). Another hypothesis regarding face perception is that it is based on the processing of configural information about a face. It has been difficult to formalize the difference between the two hypotheses, because for typical face recognition experiments, these two hypotheses make similar predictions and are thus not distinguishable from one another. The present study aims at formally differentiating between the two hypotheses through integration with a mathematical model, the General Recognition Theory (GRT). Past efforts at formalization are reviewed and their limitations are outlined. The experiment reported here overcomes a number of past shortcomings to strongly suggest that face processing is an interactive processing of face parts.

Configural and Holistic Processing

An early attempt at operationalizing face processing to the visual properties of a face has been to equate this type of processing to configural processing—processing of the configuration of features rather than the features themselves. Rhodes (1988) distinguished between first order features, which consist of readily labelled features such as the eyes, nose, and lips, and second order features, which consist of the configuration, or spatial relations, between the first order features. Rhodes then suggested that face processing involves second order information, and is thus mostly configural. To make measurements of the role of the first and second order features in face processing, the author used multidimensional scaling on similarity judgments amongst a set of faces that varied on both first and second order information. Although she observed both first and second order information as being relevant for processing, her formulation was unreliable because attributes associated with first order features, such as mouth width, could very well be associated with second order features because they also affect the configuration of the face.

Diamond and Carey (1986) distinguished between first and second order *relational* information. In their formulation of configural processing, first order relational information is used to distinguish between classes of objects, but second order relational information is thought to be needed to distinguish *within* a class of objects. Thus, in face perception, second order relational information is most important because it is concerned with perception within a specific object class. Second order information, in this formulation, is thought to consist of the relationship of the various features relative to a prototype. Furthermore, this formulation is not unique to faces but can be generalized to all object classes. Thus expertise in perception of one object class could give rise to reliance on second order relational information, and this is supported by the finding of an inversion effect (where recognition of an inverted face is much poorer than the upright face or even other inverted objects) for dog experts (Diamond & Carey, 1986).

Tanaka and Farah (1991) presented results that dispute the above formulation. They taught participants to identify two types of dot patterns, one that followed a general configuration, and another that did not. Although they observed an inversion effect for both, they did not find a difference between the two types of patterns, suggesting that processing of second order information is not necessary for the inversion effect to be observed.

More recently, Farah et al. (1998) suggested that holistic face processing can be best understood in relation to the processing of other objects: face processing involves less part-decomposition than the processing of other objects. They reported four experiments in support of their operational definition, the first of which is most pertinent to this paper. In this experiment, participants viewed pairs of faces with features that were either the same or different. Following the presentation of the faces, a word probe was presented (i.e., 'nose', 'mouth', etc.). Upon seeing the probe, participants then decided whether the two previously viewed faces were similar or different on the probed feature. Farah et al. (1998) observed that the participants' discrimination ability for the probed item depended on how many other features the faces had in common. For example, if the faces only had different noses, but the other features were the same, a 'different' judgment for a 'nose' probe was less likely than if the other face features were also different. Farah et al. (1998) suggested that if faces were subject to part-decomposition, then each part could be evaluated independent of others. Since this was not found, they concluded that faces were not as well decomposed into parts as other classes of objects. The primary implication of their work is that faces are processed as wholes, or in other words, a whole face is itself a feature and is not represented as a conglomerate of features.

This finding is important because it suggests that the other features influence the discrimination ability for any feature. Therefore, the perception of any feature is dependent on the perception of the other features. This relationship is one that is central to the present attempt at formalization of holistic and configural processing.

Formal Models

Using the above formulations, we still do not have a formal model in which we could describe holistic or configural processing in detail. Furthermore, we have a number of formulations that seem orthogonal to one another, yet descriptive of holistic face processing: an emphasis on configural information (Rhodes, 1988), another on the use of prototypes (Diamond & Carey, 1986), and yet another based on the level of part-decomposition (Farah et al., 1998). A common theme amongst all these formulations is that they describe face processing as an interactive processing of face parts. For example, configural information, or second order relational information, can be thought of as the result of an interactive processing of the different parts: the distances or sizes of one feature to another affects the distance of that feature to a third one. Farah et al.'s (1998) formulation of face processing as involving 'less-part decomposition' can be thought of as basically greater part-integration. This approach emphasizes the greater interdependence of parts in a face as compared to other objects, and this interdependence is synonymous with interactive processing. As we shall see below, the concept of interdependence is useful in formally defining holistic face processing.

Sergent (1984) formulated holistic face processing as the interactive processing of face parts. To assess this type of process, she measured reaction times in a simultaneous same-different matching task to Identikit stimuli (i.e., line drawings), as well as dissimilarity ratings. Using the three dimensions of facial contour, eyes, and internal spacing, she concluded that face processing is based on the interactive processing of face parts. Her results were supported by two analyses, one based on regression of the reaction time data, which suggested an interactive influence between internal spacing and facial contour, and another analysis based on multidimensional scaling solutions for the dissimilarity ratings, which she observed to deviate from a perfect cube. However, the study has some central flaws that prevent us from drawing robust conclusions. For example, because perception of line drawn faces is different from realistic face photographs, it is not certain how much her findings tell us about natural face perception (Leder, 1996). Also, her analyses, using regression of multidimensional scaling, may be questionable (Macho & Leder, 1998). Finally, the strongest interaction that she observed was with the dimension of internal space, which is the relative space between all features inside the face (i.e., the distance between the eyes, between the eyes, nose, mouth, eyebrows, etc.). Because this dimension is quite global in nature (any change on this dimension results in changes in all features), it prevents us from drawing any realistic, general conclusions about how face perception is done.

In another approach, Macho and Leder (1998) assessed the interactive influences of facial features on the perception of a face by formalizing holistic face processing using a mathematical model called the logit model. Before describing their model, it is best to describe the experiment. A single face photograph was digitally manipulated on three levels of the three dimensions of eye-to-eye-distance, nose width and size of mouth, to yield 27 versions of the face (i.e., (1) far eyes, narrow nose, small mouth, (2) far eyes, narrow nose, big mouth, (3) etc.). Participants were asked to categorize the 27 faces individually as being most similar to one of two test faces, each one representing manipulations on the extreme ends of the two dimensions (i.e., one having far-eyes, wide-nose, big-mouth, and the other close-eyes, narrow-nose, and small-mouth). The authors then fit a probability model to the participants' responses. The logit model assumes that the probability of selecting one test face over another is independent of the features of the target (i.e., viewed) face. This assumption implies that the features of the target do not interact to bias judgment or perception of a target face towards one of the test faces. Therefore, if the model fits their data, this suggests that interactive influences among face components do not bias perception of that face.

Macho and Leder (1998) provided support for their model, suggesting that face perception is not based on an interactive processing of face parts and is thus not holistic, because a model assuming independence in perception between the face features best fit their results. Their results are, of course, very surprising, given that there are already numerous experiments suggesting that face processing is based on holistic processing of some sort, but none had formalized what is meant by holistic to the extent that Macho and Leder (1998) had. However, their study had a number of limitations that are described below.

First, there is a simple issue of impoverished response: participants viewed 27 different versions of one face and their task was to judge how similar each face is to two of the three extreme combinations of the features. Note that the two test faces were the same throughout, which raises a simple question: why did they not use the third extreme combination of the face features? In any case, given that participants' responses were limited to the two faces, their response was impoverished, preventing them from clearly expressing the variety of their perceptual experiences. Furthermore, 25 out of 27 times the participants *had* to be wrong, and this greatly restricts our understanding of the participant's performance.

Second, there is a central issue of sensitivity. If the participants' responses were impoverished, then perhaps there were perceptual interactions of the features, but that they could not be assessed using this methodology/model. Consequently, because the logit model was fit to such an experimental design, the model could be inherently limited in assessing interactive influences of the feature combinations.

Third, the researchers assumed that decisional processes involved in the judgment task were not influenced by the different feature combinations. In other words, the assumption was that the decisional processes were unaffected by feature combinations. Such behaviour is better tested than assumed, and thus a model is needed that can account for and measure such behaviour rather than assume it invariant.

Finally, from the previous critiques of Macho and Leder (1998) and the earlier discussion regarding Experiment 1 in Farah et al. (1998), it can be observed that in our search for a formal definition of holistic processing in perception we are talking about a number of different interactive effects or dependencies between various features:

(1) Interactivity between face parts within a face, or the extent and direction to which perception of one face feature may depend on perception of other features for a given stimulus.

(2) Experiment 1 in Farah et al. (1998) shows that discrimination on one feature between two faces is influenced by the number of other features the faces have in common. In other words, discrimination ability for one feature is dependent on the other features present. This is an issue of perceptual inseparability.

(3) Given that decisional boundaries could very well shift for different feature combinations, there is a possibility that decisional processes are influenced by the combination of features present. We shall call this decisional dependence.

We can see at once that the logit model of Macho and Leder (1998) ignores the latter two dependencies and focuses only on assessing perceptual dependencies using a model that is too insensitive to even assess that type of dependence. Clearly, a more rigid formulation of these dependencies is needed within a unified framework to effectively describe the processes involved in face perception and discrimination. Does such a framework exist? Yes.

The General Recognition Theory

The General Recognition Theory (GRT; Ashby & Townsend, 1986) extends the theory of signal detection to multiple dimensions. Signal Detection Theory (SDT) is limited to separating decisional and perceptual effects for detection along only one dimension. When multiple dimensions are involved, a number of different interactions could arise that need to be separated from one another. In this more complex case, one must first be able to separate out decisional factors from perceptual ones, but it must also be noted (as we have done so above) that more than one type of perceptual factor may be at work: one relating to discrimination ability and another to perceptual dependences (i.e., perception based on perceptual interaction of face parts).

The GRT was developed by Ashby and Townsend (1986) in response to various issues that had been raised in perception research. These issues concerned the notion of independence and separability of the perception of stimulus dimensions as well as decisional factors. The GRT is a formal method for assessing independence and separability in terms of both stimulus dimensions and decisional processes.

Multidimensional Signal Detection Analysis (MSDA) was developed by Kadlec and Townsend (1992a, 1992b) to facilitate the implementation of GRT in perception studies. This analytic method maps the traditional SDT parameters of sensitivity (d') and response bias onto a multidimensional scheme and permits us to analyze both interaction and independence between various stimulus dimensions.

MSDA was specifically designed to analyze the effects of manipulation on the perception of a stimulus when the manipulations are varied on a number of dimensions (Kadlec, 1995). An example study would be one that looks at the dependence of the perception of eye distance and nose length. First, one would need a feature-complete factorial design, which may be created by manipulating each of the two dimensions on two levels. Thus the eye distance could be varied on two levels (eyes close and eyes far) with the same manipulation being used for nose length. Table 1 demonstrates this matrix for all variations of stimulus A.

In a typical experiment employing this design, participants view the set of stimuli, one at a time, varied on the given dimensions, and make judgments about their location at the different levels of the dimensions. In other words, a participant who views A_{bj} must try to categorize this stimulus on both dimensions. The correct response for this stimulus is to identify it as one having eyes far apart and a long nose. There are three types of errors that can be made in the categorization of each stimulus in such a 2 x 2 design. These errors can be tabulated, resulting in proportions of each type of error, which then represents the volume under the distribution in one of four possible response regions.

These response regions can be represented by Gaussian distributions in multiple dimensions – in this case, a 2 dimensional Gaussian distribution (see Figure 1). The assumption here is that the perception of a stimulus is not absolute and that each time it is presented, the perception of the stimulus may be slightly different from the last. However, most of the time the perceptions of a given stimuli will be the same, and so if each percept of the stimulus was represented as a dot in a "perception space", then these dots would cluster around an average percept forming a distribution. The GRT makes the assumption that the distributions of percepts in this "perception space" are Gaussian.

How could we, from these data and types of distributions, answer questions regarding independence of dimensions and decisional boundaries? In order to answer such questions, a slightly different view of the graphs must be utilized. Consider a plane passing horizontally through all the normal distributions in this multidimensional space at a given density level. Examining such a plane from above would yield a topography of the distributions, as illustrated in Figure 2.

The location and orientation of these distributions, along with the orientation of the decisional boundaries, reveal information regarding perceptual, discrimination, and decisional dependence or independence. Perceptual independence (PI), in the language of GRT, is thought to hold when for a given stimulus A_{bj},

$$F_{bj}(x,y) = g_{bj}(x)*g_{bj}(y)$$

In this equation, g(x) refers to marginal densities, which are obtained by integrating (measuring the area under the curve) the two-dimensional density distributions across one dimension. Marginal densities can be thought of as the picture of a density distribution as would be taken from having a camera parallel to a dimensional axis.

Perceptual independence is a strictly statistical form of independence and can be likened to coin toss probabilities, where the probability of obtaining two heads (assuming a fair coin, p(H) = 0.5) is equal to the product of the probabilities of obtaining one head by the same probability, or:

$$p(2H) = p(1H)*p(1H) = 0.25$$

Thus PI asserts that the perception of one dimension within a stimulus is not dependent on the perception of the other dimension. From the topographical diagram, PI is represented by circular distributions and distributions that are elliptical but parallel to the axis of one dimension. From the diagram it may be observed that within stimuli 1, 3, and 4, the two dimensions are perceptually independent and the two-dimensional density distributions are equal to the product of the marginal densities of the stimulus on the individual dimensions.

One way of conceptualizing perceptual dependence is to think of it as indicative of an illusion of some sort—if the perception of one stimulus dimension affects the perception of the other stimulus dimension, then the resulting percept is an illusion created from that specific combination of dimensions. Thus the tilted ellipse 2 in Figure 2 suggests that this stimulus was perceived as displaying even longer nose and longer eye distances than was present. This observation was made possible by the fact that the errors associated with this stimulus were unevenly distributed.

Another GRT concept relating to perceptual interactions concerns perceptual separation (PS), which is taken to hold true when within one level of one dimension, the levels of the other dimension do not affect perception. In this case,

$$g_{i1}(x) = g_{i2}(x) \text{ and } g_{j1}(x) = g_{j2}(x)$$

where 1 and 2 refer to stimuli 1 and 2, i and j represent the two levels of eyebrow curvature, and g(x) refers to their marginal density distributions. If the two marginal distributions at the levels of one dimension (say, long nose) are equal, then levels of the other dimension (eye distance) do not influence the perception of the eyebrow curvature.

A third form of independence is decisional separability (DS). Within the unidimensional signal detection framework, a decisional boundary was set between the two distributions. Similarly, within the multidimensional framework, some decisional boundary must be set. This decisional boundary is set by the participant and defines the area within which a stimulus will be identified by its specific characteristics. In our example, a decisional boundary must be set in order to differentiate between faces that vary differentially on eye distance and nose length. In other words, the decisional boundaries divide the multidimensional space into regions that define the pattern of response to the stimuli. Thus the measure of DS ensures that decisional factors are separated from perceptual ones, and that their interaction with stimulus properties are separately accounted for. Within this context, DS is observed when the decision about one dimension is not influenced by the decision made on the other dimensions. In our topography (Figure 2), DS holds when the decisional boundaries are parallel to the dimensional axis.

It can be noted that GRT overcomes a number of the limitations inherent in Macho and Leder (1998) logit model: (1) it allows for a rich response set, (2) it is capable of integrating the rich response set, (3) it makes no assumptions about the influence of decisional factors, and (4) it assesses an exhaustive set of possible perceptual and decisional interactions that may occur amongst face parts. Therefore, it is an ideal model for describing the nature of holistic and/or configural face representations. In fact, this methodology has previously been applied to face perception by Thomas (2001), but as we shall see below, her study does not conclusively tell us whether or not GRT is an adequate framework for formally defining holistic face processing.

GRT and Face Perception

Thomas (2001) employed GRT in two experiments involving feature-complete factorial designs of different face parts. She made use of the three dimensions of (1) eye-to-eye distance, (2) mouth width, and (3) nose length. These dimensions were realized in manipulations made to a line-drawn, abstract face. In other words, she did not use a line-drawing algorithm to create her faces, but instead used an artificial line-drawn face, as shown in Figure 3.

Of the four participants in her studies, two were tested on the combinations of dimensions (1) and (2), and two were tested on combinations of dimensions (1) and (3); each dimension was manipulated on two levels. In Experiment 1, participants had to make speeded classifications, whereas in Experiment 2, participants were given adequate time to make their judgements. In both cases, participants viewed the target faces for 125 milliseconds. Results from Experiment 1 suggested that the dimensions of mouth width and nose length are both perceptually separable from the dimensions of eye-to-eye distance, because reaction times for classification of the mouth width and nose length were unaffected by changes in eye-to-eye distance, and vice versa. However, Experiment 2 suggested that a perceptual dependence does exist between these different dimensions (a negative correlation was observed for eye distance and nose length in two of four subjects), which is consistent with an interactive processing model of holistic face perception. This interaction was found to be greater for parts that are geometrically closer to one another—i.e., a greater perceptual dependence was observed between nose length and eye-to-eye distance than was between mouth width and eye-to-eye distance. Although Thomas (2001) does not readily provide an interpretation of why this happens, one possible interpretation is by analogy to the Gestalt law of proximity, which dictates that objects that are closer together are perceived as single objects. In this case, it could be that proximal face parts, such as eyes and nose, display greater perceptual dependence because it is more likely that those proximal items will form local Gestalts themselves. In contrast, mouth and eyes, being more distal, are less likely to form Gestalts, and therefore show a weaker degree of perceptual dependence.

However, at least two other lines of research call into question these results. First, recall Experiment 1 in Farah et al. (1998). In that study, participants simultaneously viewed two faces that varied on the number of features they shared in common. After the presentation of the two faces, participants viewed a feature-probe word (i.e., *nose*) and had to respond whether the two faces shared the same probed feature or not. The researchers observed that the degree of errors people committed on this task was dependent on how many features differed between the two faces. In other words, the ability to discriminate between two features varied as a function of other features in the faces. In GRT terms, this would be equivalent to the lack of perceptual separability (i.e., perceptual inseparability). This was not observed in any of the four subjects in (Thomas et al., 2001).

Second, Thomas (2001) only tested four subjects, and observed significant interactions in only 2 of those 4 subjects. Therefore, her study has very little power to reject the hypothesis of non-interaction, even though she did so. A more important limitation in Thomas' study is that she did not elaborate on the concepts of GRT and how they relate to face perception; I hope to bridge this gap below.

Why did Thomas (2001) get such unexpected results? Recall that not only did Thomas (2001) employ line-drawn faces, but also that those line-drawings were of artificial faces. Such stimuli are very uncommon in face perception studies because they bare very little similarity to real faces, given that they are void of any three-dimensional information, texture, shade, colour, and are strongly impoverished in their representation of face features. Given that perception of line-drawn faces extracted from realistic photographs is already shown to be different from face perception based on the photographs themselves (Leder, 1996), it is uncertain how much ecological validity an *artificial* line-drawn face would have. Also, Thomas (2001) did not compare the interaction of the face parts of upright faces to inverted faces to show that such perceptual dependencies (as measured by GRT) only occur for the upright but not inverted faces, which contain the same features and configuration as upright faces, but should not be processed holistically. Such a comparison is central if one is to conclude that GRT is capable of formalizing holistic face perception as interactive processing of face parts.

The present experiment overcomes the limitations inherent in the previous attempts by Macho and Leder (1998) and Thomas (2001) by (1) using the GRT to assess the perceptual and the decisional interactions of face parts, (2) using face photographs as opposed to line-drawn artificial faces (see Figure 4), (3) employing a comparison condition using inverted faces to test whether interactive processing is unique to upright faces, and (4) testing a larger group and analyzing across individuals for a more generalizeable analysis. The question asked is whether face processing (and by extension, whether configural and/or holistic face processing) is the interactive processing of face parts. If so, then one would expect to observe interactions for upright faces but not for inverted faces. Within the context of GRT, it is predicted that both perceptual dependence and inseparability will be observed for upright faces, but that they will not exist for inverted 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 stimuli 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 4.

Apparatus

The experiment was presented on a Macintosh computer, using the Psychophysics toolbox for Matlab (Brainard, 1997; Pelli, 1997). Participants were typically 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 dim-lighted environment.

Procedure

Participants were informed of the nature of the task. They 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 centre 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. A half-second delay followed the target face, and subsequently the test faces appeared. Figures 5 and 6 summarize the procedures employed in this study.

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 judgement 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, to walk around, relax their eyes, etc. 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.

Using MSDA-2 (Kadlec, 1995), the data were subjected to multidimensional signal detection analyses to make estimates of the different types of interactions, namely Perceptual Separability, Perceptual Independence, and Decisional Separability. All tests were two-tailed Z-tests, with $\alpha = 0.05$. Tables 2 and 3 present the d' and C estimates in the macroanalyses (pertaining to Perceptual and Decisional Separabilities), while Figures 7 and 8 summarize information from the microanalyses (pertaining to Perceptual

Independence) along with the results of the macroanalyses in equal density contour diagrams.

The reader is referred to Kadlec (1995) and Kadlec and Townsend (1992a; 1992b) for full details of the analysis. Briefly, the analytic method involves a macroanalysis, which reveals information about perceptual and decisional separability, and a microanalysis, which reveals information about perceptual dependence. 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—if values are significantly different from one another, this suggests a violation of separability, and the direction of this interaction can readily be ascertained by looking at the d' and C estimates.

In the microanalysis, conditional d's are calculated. For example, for stimuli with short noses, d' for eye distance judgements are calculated when such stimuli are appropriately identified as having short noses, and also a d' conditional on when the component of short nose was not properly identified (subjects confused the face with one of long nose) is calculated. As a result, at each level of one dimension, two d's can be estimated, which are then used to calculate the tilt of the ellipses. If, for example, items with short noses yield larger eye distance d's when they are properly perceived as having short noses than when they are perceived as having long noses, then this would result in at least one of the ellipses to be tilted, with the side with the tilt being away from the level of long nose. One can think of these conditional d's as distances between the corresponding sides of two distributions representing stimuli at one level of one dimension (i.e., the distance between the left side of distributions corresponding to the stimuli with A_{ai} and A_{aj}, vs. the distance between the right sides of those distributions. The difference between these distances is used to estimate the tilt of the ellipses).

Tables 2 and 3 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 is evidenced by the significant differences observed in the d' estimates of eye distance across nose length and also nose length across eye distance, an effect that is only observed for upright faces but not inverted faces. Figures 7 and 8 summarize the results for perceptual and decisional separability, and also represent information for perceptual dependence. It can be seen that more ellipses are tilted for upright faces than for inverted, suggesting greater perceptual dependence for upright faces than inverted ones.

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 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 decisional effects were observed for inverted faces.

Finally, as noted above, perceptual independence failed on several occasions in upright faces, but failed less so for inverted faces. This type of interaction can best be thought of as an association or a Gestalt effect. These results indicate 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.

Discussion

This study attempted to assess whether faces are processed as an interaction of their components. Participants viewed faces that varied on two dimensions and made identification judgements for the faces. The pattern of their errors was used to make estimates of interactivity using GRT as the integrative model. It was shown that a number of such interactions occur for upright faces, but not inverted faces, suggesting that face processing does involve the interactive processing of face parts.

Relationship to Previous Findings

The present results expand on previous experiments that employed formal definitions of holistic and configural processing. Prior to a discussion of definitions of holistic and configural processing, differences between the present findings and previous studies published on the topic will be explored.

As discussed previously, Macho and Leder (1998) utilized the logit model to assess interaction of face dimensions, but they did not observe any such effects. It is important to note exactly which types of interactions their model was sensitive to; specifically, their model was insensitive to presence/absence of decisional and perceptual separability, and would have only been able to detect the presence/absence of perceptual dependence. Why were no interactive perceptual effects observed? Perhaps because of reasons outlined above—the model was insensitive, and participants' responses were impoverished, amongst other limitations (see above). The present findings are at direct odds with Macho and Leder (1998) in that all three types of interactions (perceptual inseparability, decisional inseparability, and perceptual dependence) were observed in the present study (note that the logit model assumes decisional separability but does not test it; thus Macho and Leder (1998) never actually assessed this type of interaction). Given that the present study utilizes a more sensitive model that is sensitive to all three types of interactions, and that the methodology does not impoverish the participants' response set, it is more appropriate for the assessment of interactive face processing.

Thomas (2001) also used the General Recognition Theory to analyse data in a similar study of face perception. However, her study was limited in three important aspects. First, her stimuli lacked a great deal of similarity to real faces, as is evident in Figure 3. This is an issue of validity—perhaps her stimuli were tapping into processes that are not necessarily related to face processing. However, because she did observe perceptual dependence between the dimensions of eye distance and nose length, it is perhaps not entirely the case that her stimuli were invalid. Yet, she did not observe perceptual inseparability of the two face dimensions, and given that much of the previous results (e.g., Farah et al., 1998) would give one reason to expect perceptual inseparability of the two dimensions, this is indicative that her stimuli may have been invalid.

Finally, Thomas (2001) tested only four participants, and only two of them were tested on the dimensions of eye distance and nose length (which, incidentally, was the only condition for which she observed interactions). It is unclear why she only used four subjects, since such small subject number limits the generalizability of her study. In the present experiment, 47 subjects were tested to address this limitation.

Perhaps the most substantial shortcoming of past studies employing mathematical models is that they have failed to use proper controls—the studies only looked at perception of upright faces, but did not compare it to perception of other objects such as inverted faces. As a result, any conclusions drawn from these studies are inconclusive, because there is no proof that the interactions they observed (or did not observe) are due to face perception *per se*, but may have been due to other confounding variables. The present study overcomes this important shortcoming by employing proper controls through inclusion of a condition where inverted faces were presented. The results observed for upright faces (i.e., the interactions) can therefore be properly attributed to *face perception*, because the same pattern of interactions were not observed for the inverted faces.

In summary, the present study has demonstrated that (a) face perception involves interactive processing of face parts, (b) that these interactions are specific to faces and not to other objects such as inverted faces, and (c) the interactions include stimulus-specific interaction of face parts as well as cross-stimuli interactions, along with decisional biases.

Separable Configural and Holistic Effects

As previously noted, two general hypotheses have been proposed that describe face processing as either a holistic (i.e., Tanaka & Farah, 1993) or configural (i.e., Diamond & Carey, 1986) process. Although differences between the two definitions have recently become more muddied (for example, Farah, 1994, proposes a definition of holistic processing whereby relations between parts are more important than the parts themselves—a definition that is hard to distinguish from a configural definition), it is important to explore whether the two definitions are truly addressing the same phenomenon. Is configural processing just holistic processing? This is the main question.

Holistic Face Representation

To present an answer, we must look first at the extremes of each definition. At its extreme, a holistic processing account would suggest that the features of a face are not decomposed and so the representation of a face does not separately identify a nose, a mouth, etc., but that the face is represented as an undivided whole. A simpler way to consider a holistic face representation is to think of it as the entire face being a single feature—if the representation of a face is not decomposed into its parts (as proposed by Farah et al., 1998), then the whole face must be treated as a single feature. Thus, at an extreme, a holistic face representation is a form of representation where the whole face is treated as a single feature.

If the whole face is treated as a single feature, then our brains must have such feature detectors. Neurophysiological recordings suggest that this is the case (e.g., Desimone, Albright, Gross, & Bruce, 1984), adding weight to the holistic hypothesis. What would a simple form of such whole face detectors be, especially in the context of the present study? In the context of this study, a whole-face detector would exist for each one of the four versions of the face (since each face must have its own representation), and for each face, processing of one aspect of the face (i.e., eye distance) would depend directly on another aspect of the face (i.e., nose length), such that if, for a given face, the eyes are close together, this directly affects how the nose is perceived.

It can now be readily seen how well such a notion fits within the GRT framework—a holistic face representation could be identified to exist for a face if there is a detectable perceptual dependence *for that face*. Recall that perceptual dependence, as defined by GRT, is a within-stimulus type of interaction whereby for a given stimulus, perception of one component directly affects perception of another component. This is perfectly compatible with the definition of holistic representation that we derived above. Thus, a holistic representation can be said to exist for a face if perceptual dependence amongst its components is detected.

This is exactly the kind of interaction that Thomas (2001)was able to observe (for all four stimuli in her study), but that is only present strongly in two of the stimuli in the present study. For reasons pertaining to limitations of her study, as previously outlined, the discussion will be restricted to the present data and its relationship to holistic representation. In the present study, perceptual dependence was observed for two of the stimuli—eyes close and long nose, and eyes far and short nose. In fact, the nature of the interaction observed directly relates these two stimuli together; as Ashby (1989) and Kadlec (1999, personal communication) propose, when perceptual independence fails, the tilt of the distributions can be best interpreted as an emergent property and discussed in terms of associations or correlations between the different percepts. Thus my results indicate an inverse association between nose length and eye distance—as eyes are perceived as being further apart, the nose is perceived as being shorter, whereas as the eyes are closer together, the nose is perceived to be longer; it should be noted that the relationship between eye distance and nose length is bi-directional, so it could equally well be said that as the nose is perceived to be longer, the eyes are perceived to be closer together, etc.

This is an intriguing finding, because it relates directly to the physical studies of face structure as discussed by Enlow and Hans (1996), whereby faces can be globally classified into two categories—Dolichocephalic faces have long noses and eyes that are close together, while Brachycephalic faces will have short noses and eyes that are far apart (see Figure 9). The present results indicate that we form holistic representations that are consistent with this physical characteristic of faces. Thus, the present results suggest that our whole-face detectors are tuned to the physical differences between faces as evident by the inverse association of nose length and eye-to-eye distance which are, in fact, often united dimensions in face variability.

Configural Face Processing

From the above discussion, it can be inferred that a holistic face representation is a valid concept, and that the present results support the notion of holistic face representations and also suggest that the nature or pattern of such a representation is quite closely related to the physical characteristics of a face. But perceptual dependence, an identifier of holistic representation, was not the only interaction observed in the present study. Perceptual separability, a more global form of perceptual interaction, also failed for both dimensions of eye distance and nose length. Specifically, the data suggest that discrimination of eye distance is poorer when the nose is short, and that the discrimination of nose length is more difficult when the eyes are far apart. This effect is independent of perceptual dependence, which is related to holistic face processing. How can we explain the perceptual inseparabilities observed in the present study? What concept can we use to discuss this effect?

Perhaps a modified definition of configural processing would best describe the observed effect. In its extreme form, a configural coding hypothesis would imply that face components are decomposed, and that the relationship between them is represented and not the components (i.e., features) themselves. Thus face features are decomposed, but their relation to one another is encoded (Diamond & Carey, 1986).

The main prediction of the configural hypothesis is that changes in the configuration of a face will affect its processing (Diamond & Carey, 1986; Rhodes, 1988), and there is good evidence to support this prediction (e.g., Tanaka & Sengco, 1997). Tanaka and Sengco (1997) presented participants with a recognition task whereby in the test images, the distances between the eyes were changed. This change was detrimental to recognition ability, suggesting that configural changes had disrupted the holistic face representation by influencing how other features were recalled. This is analogous to Farah et al. (1998), who showed that discrimination ability for features was influenced by other features, again suggesting that one part of a face may influence perception and subsequent discrimination of another part of the face.

But this final issue pertaining to the configural processing of a face does not really involve the encoding of a configuration *per se*, since the studies by Tanaka and Sengco (1997) and Farah et al. (1998) do not preclude the possibility that configural information is actually encoded; in fact, these researchers proposed that their results support a holistic account of face perception and not a configural one. This is evidence of the difficulty inherent in working with vague notions and concepts of configural and holistic types of representations, since it can be very difficult to support one vague concept without simultaneously supporting the other as well. However, within the context of GRT some clarity may be afforded.

Already a formal definition of a holistic representation has been given (see above) in the context of GRT—a holistic representation is reflected in the presence of perceptual dependencies among different components of a face, *within a stimulus*. Also noted previously, results from Farah et al. (1998) and Tanaka and Sengco (1997) suggest the absence of perceptual separability between facial components (i.e., perceptual integrality amongst face components). Thus it was surprising that Thomas (2001) did not observe a perceptual inseparability. The low subject number and her choice of stimuli, which may have limited the power of her study, might explain this. Another possibility is that perhaps her stimuli did not tap into configural processing (Leder, 1996), and it is perhaps exactly this type of processing that results in perceptual inseparability effects.

The definition of configural processing proposed here implies that, as consistent with traditional notions of configural processing, face components are decomposed and processed by separate 'channels' of processing (Ashby, 2000; see figure 10). However, as evidenced by the present data, there is much cross-talk between the channels. As a result, processing in one channel is affected by another, resulting in noise and detrimental discrimination performance. Although one may be tempted to simply pass these results as also supporting a holistic model of face representation, it should be made clear that these are not due to Gestalt effects for the simple reason that a Gestalt account would not permit for any decomposition of face parts.

It may be possible to discuss the results from another perspective. It could be said that there are no processing channels involved, but instead, detectors exist that are sensitive to specific configurations—when visual information falls properly on the receptive fields of these configuration detectors, they fire optimally, enabling better discrimination. But when visual information does not properly fit their receptive fields, the firing of these detectors is reduced, resulting in reduced discrimination performance. This may explain why discrimination of eye distance is easier when the nose is long than when the nose is short—presumably because the configuration detector is optimally tuned to longer noses.

In either of the formulations, the effect is configural in nature, and corresponds to the GRT concept of perceptual inseparability. Thus, perceptual inseparability, as estimated using the GRT technology, can be used as an estimate of configural processing, whereas perceptual dependence can be used as an estimate of holistic processing. In summary, within the context of GRT, concepts of holistic and configural processing can be integrated and clearly and separately defined, giving us new power to investigate these distinct processes.

Computational Utility of Interactive Processing and Future Directions

The main aim of the present thesis is to propose that face processing differs from non-face object processing because the former involves interactive processing of components whereby one component affects processing of another, while the latter does not. Since faces constitute a socially vital set of objects, and since humans are quite capable at discriminating between the large variations in faces, it may be speculated that interactive processing somehow brings about this awesome capability. The intent of this section is to illustrate the computational utility of interactive processing of components in intra-class object processing (specifically, face processing).

For this illustration, consider a set of objects that differ on only two dimensions, x and y. Now, consider a normal distribution of objects within this two dimensional space, such that a larger number of member objects are clustered together around a mean, while fewer numbers of members lie in the skirts (see Figure 11). Naturally, faces that are close to each other in this space are more similar to one another and will thus often be mistaken, whereas faces further apart are more dissimilar and will be easier to distinguish.

If we make the safe assumption that the majority of faces that one knows are more or less average faces, then how can we distinguish between so many similar faces? The solution proposed here is that interactivity of dimensions exists, whereby a mental dimension is added to the face-space, and points on this dimension are calculated as a function of the other dimensions. This gives rise to a surface, as illustrated by Figure 12. This added dimension expands the distance between two similar faces, thus facilitating discrimination of similar, average faces. Thus, an interactivity account of face processing can be computationally useful because it allows one to better discriminate between similar faces by scoring them on a new, derived dimension along with their physical dimensions.

Formalization of holistic and configural processing mechanisms, as proposed by this thesis, allows for a number of future research directions. One direction is to identify the interaction between other key dimensions in face variability, as estimated by employing Principle Components Analysis (Turk & Pentland, 1991; Valentine, 1991). This statistical method identifies main components of variation amongst a set of items (in our case, faces). Once these components are identified, the GRT methodology can be used to assess if and what interactions exist between these estimated physical dimensions.

Another line of research might be to employ the present methodology to investigate memory effects related to holistic and configural representations. For example, at a perceptual memory level (Magnussen, 2000), one question is whether increased delays between presentation and test result in greater or less interaction of components. Attention effects related to configural and holistic processing can also be used to investigate whether attention, proposed to be required for feature binding (Treisman & Schmidt, 1982) is essential for interactive processing of faces. A simple study could employ the same methodology presented here, with an added divided-attention condition during presentation.

In short, through this level of formalization, a number of future directions are opened. This line of work will better enable us to decipher the mechanisms of face processing.

Summary and Conclusions

In this study, I have shown that face perception involves interactive processing of face parts. This interactivity was limited to upright faces and not inverted faces, suggesting that this phenomenon is unique to normal, upright faces. It was possible to dissociate three different interactions, two perceptual interactions and one decisional interaction. These two forms of perceptual interactions were then discussed in the context of contemporary, competing hypotheses, one suggesting face processing to be a holistic, Gestalt process and another suggesting that face processing involves mainly the processing of configuration of face parts. It was shown that the two hypotheses could be integrated within a larger framework. In the context of GRT, Gestalt processing was defined as perceptual dependence, while configural processing was defined as perceptual inseparability. The computational utility of such interactive processing of face parts was described to show how interactive processing may improve face discrimination for similar faces.

References

Ashby, F. G. (1989). Stochastic general recognition theory. In D. V. a. P. L. Smith (Ed.), Human information processing: Measures, mechanisms, and models (pp. 435-457). Amsterdam: Elsevier.

Ashby, F. G. (2000). A stochastic version of general recognition theory. Journal of Mathematical Psychology, 44(2), 310-329.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. Psychological Review, 93(2), 154-79.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433-6.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience, 4(8), 2051-62.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. Journal of Experimental Psycholy: General, 115(2), 107-17.

Enlow, D. H., & Hans, M. G. (1996). Essentials of facial growth. Philadelphia: Saunders.

Farah, M. J. (1994). Specialization within visual object recognition: Clues from prosopagnosia and alexia. In M. J. R. G. Farah (Ed.), The neuropsychology of high-level vision: Collected tutorial essays. Carnegie Mellon symposia on cognition (pp. 133-146). Hillsdale, NJ ; Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; Lawrence Erlbaum Associates, Inc.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? Psychological Review, 105(3), 482-98.

Kadlec, H. (1995). Multidimensional signal detection analyses (MSDA) for testing separability and independence: A Pascal program. Behavior Research Methods, Instruments, & Computers, 27(4), 442-458.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. Journal of Mathematical Psychology, 36(3), 325-374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), Multidimensional models of perception and cognition. Scientific psychology series (pp. 181-227). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Leder, H. (1996). Line drawings of faces reduce configural processing. Perception, 25(3), 355-66.

Macho, S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. Journal of Experimental Psycholy: Human Perception and Performance, 24(5), 1486-500.

Magnussen, S. (2000). Low-level memory processes in vision. Trends in Neurosciences, 23(6), 247-251.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4), 437-42.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. Perception, 17(1), 43-63.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. British Journal of Psychology, 75(Pt 2), 221-42.

Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: testing a theory of face perception. Perception and Psychophysics, 50(4), 367-72.

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

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. Memory and Cognition, 25(5), 583-92.

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. Perception & Psychophysics, 63(4), 625-650.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. Cognitive Psychology, 14(1), 107-141.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscienc, 3(1), 71-86.

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(2), 161-204.

Tables and Figures

Table 1

Example of a feature-complete factorial design

Nose Length	Eye Distance	
	Short	Long
Short	A_{ai}	A_{bi}
Long	A_{aj}	A_{bj}

Table 2
d' and C estimates for the dimension of nose length across eye distance for upright and inverted faces

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 3
d' and C estimates for the dimension of eye distance across nose length for upright and inverted faces

	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

Figure Caption

Figure 1. Normal probability distributions for a 2 X 2 design of eye distance and nose length

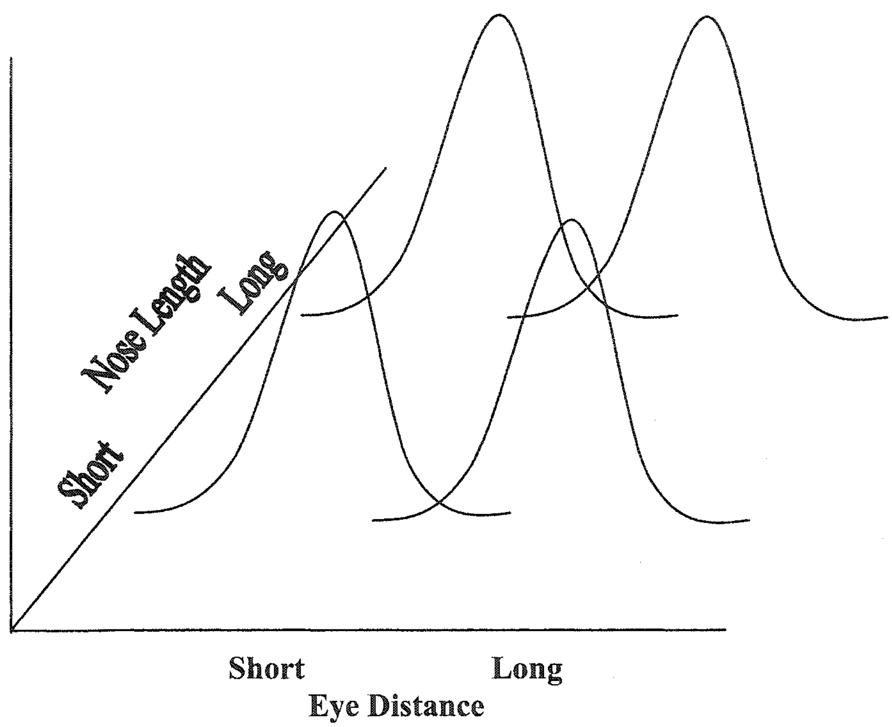

Figure Caption

Figure 2. Equal density contours for a 2 X 2 design of eye distance and nose length

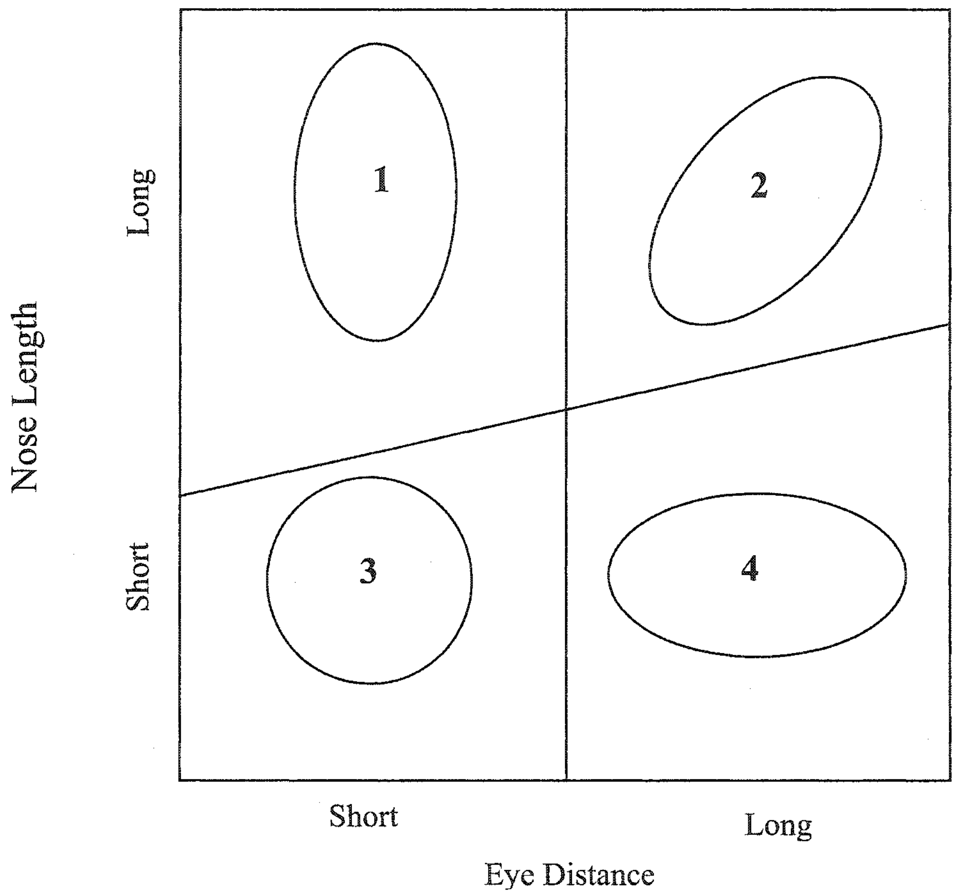

Figure 3. Example of stimuli used in Thomas (2001).

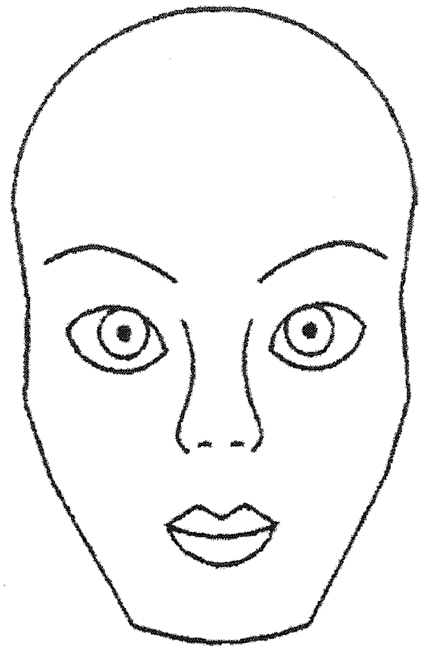

Figure 4. Example of stimuli used in this study

Figure 5. Experimental procedure in the Upright Face condition

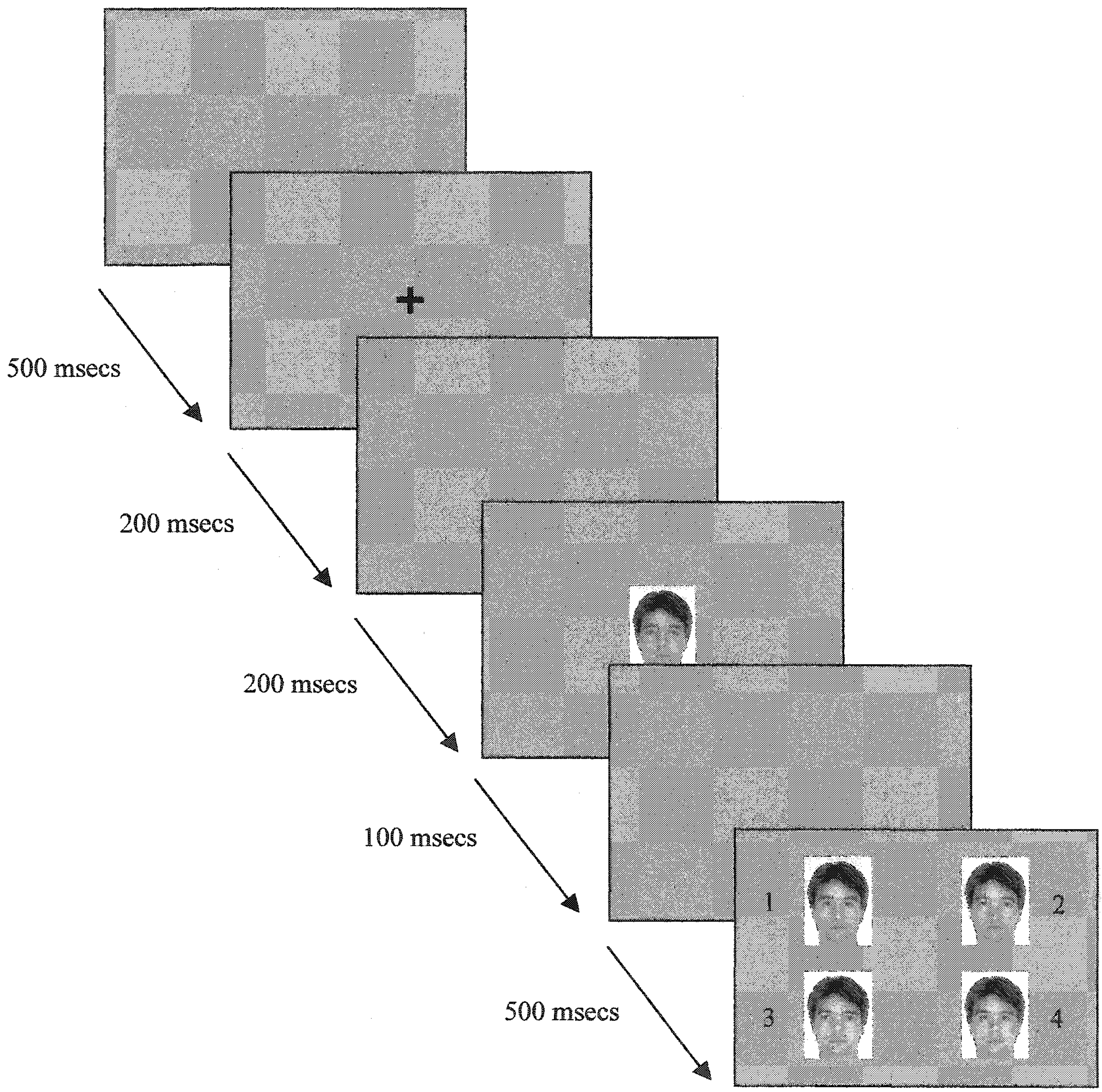

Figure 6. Experimental procedure in the Inverted Face condition

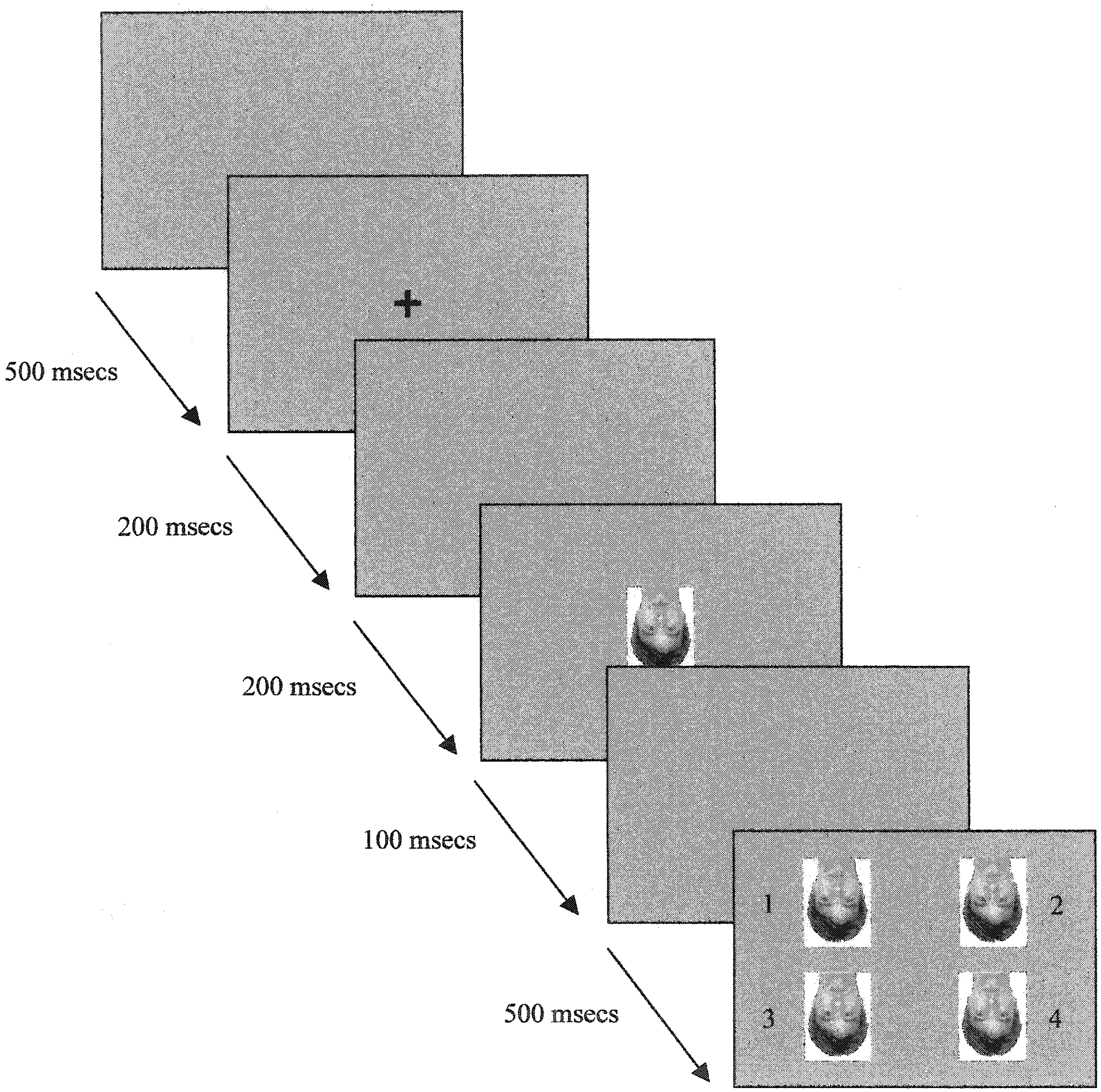

Figure 7. Equal density contour diagram for results in the Upright Face condition

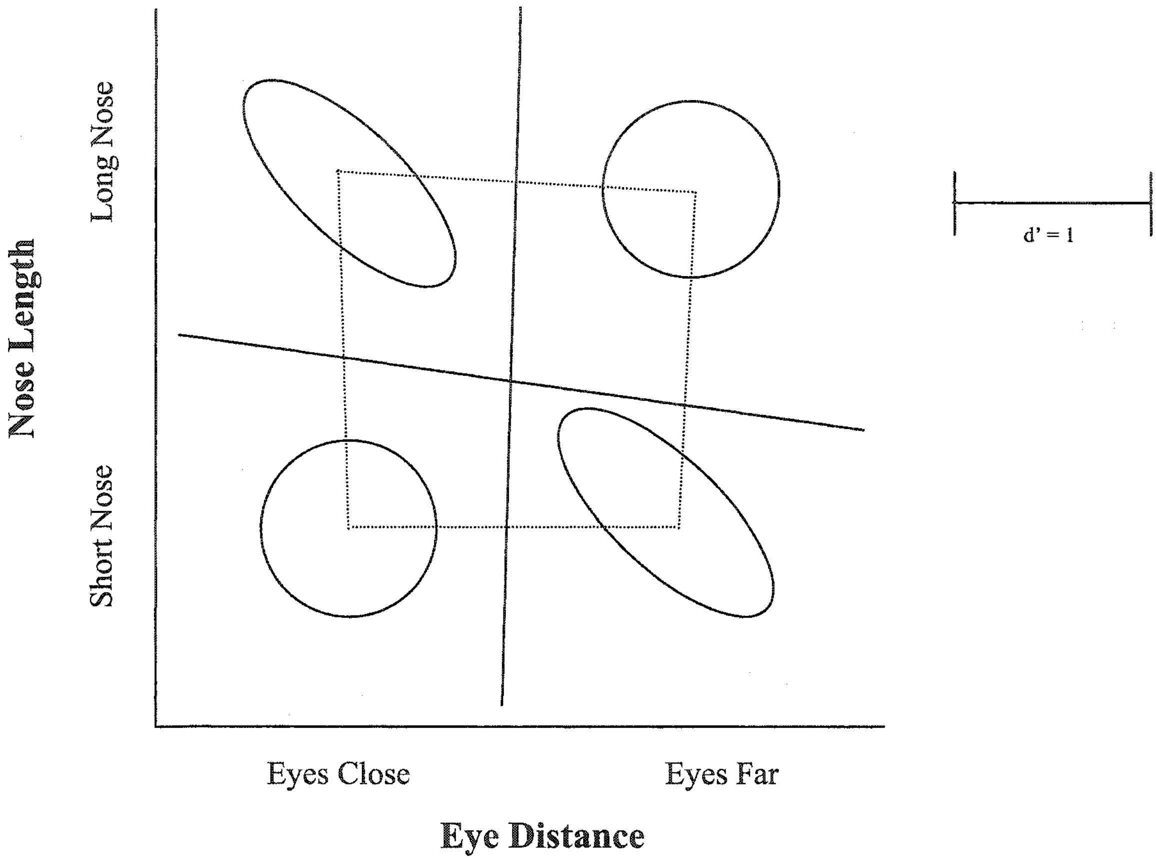

Figure 8. Equal density contour diagram for the results of the Inverted Face condition

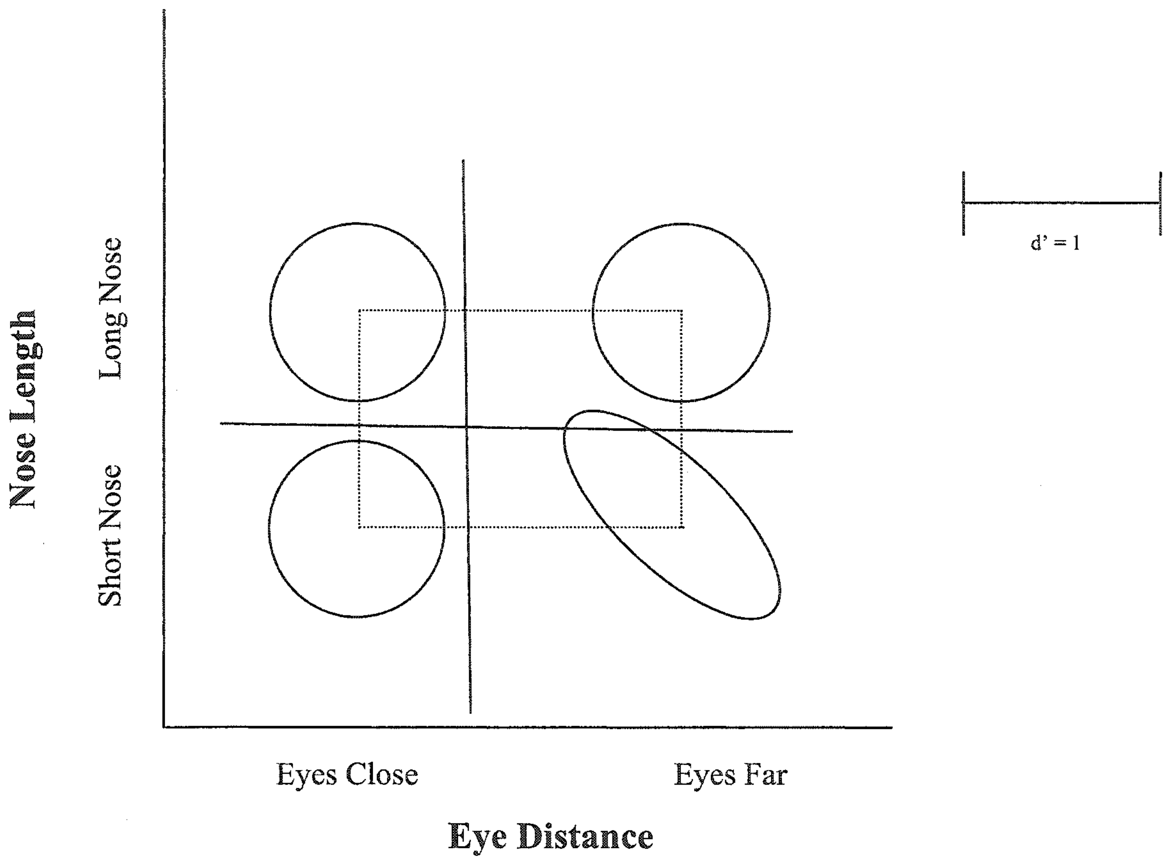

Figure 9. Dolicocephalic (top) and brachycephalic (bottom) heads produced as though the head forms were squeezed inwards (top) or stretched outwards (bottom). (Enlow, 1982)

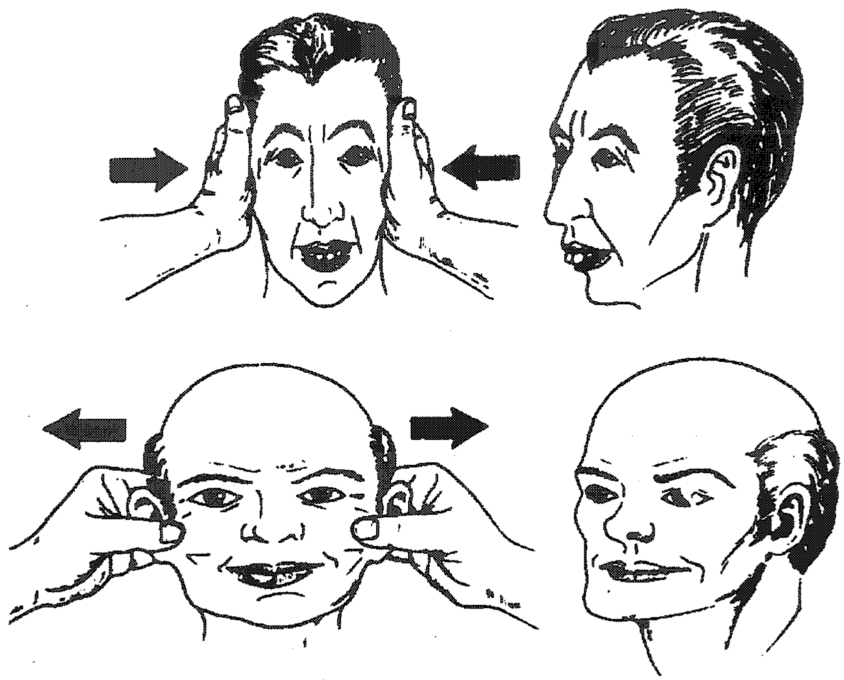

Figure 10. Schematic diagram of face component processing in the context of GRT

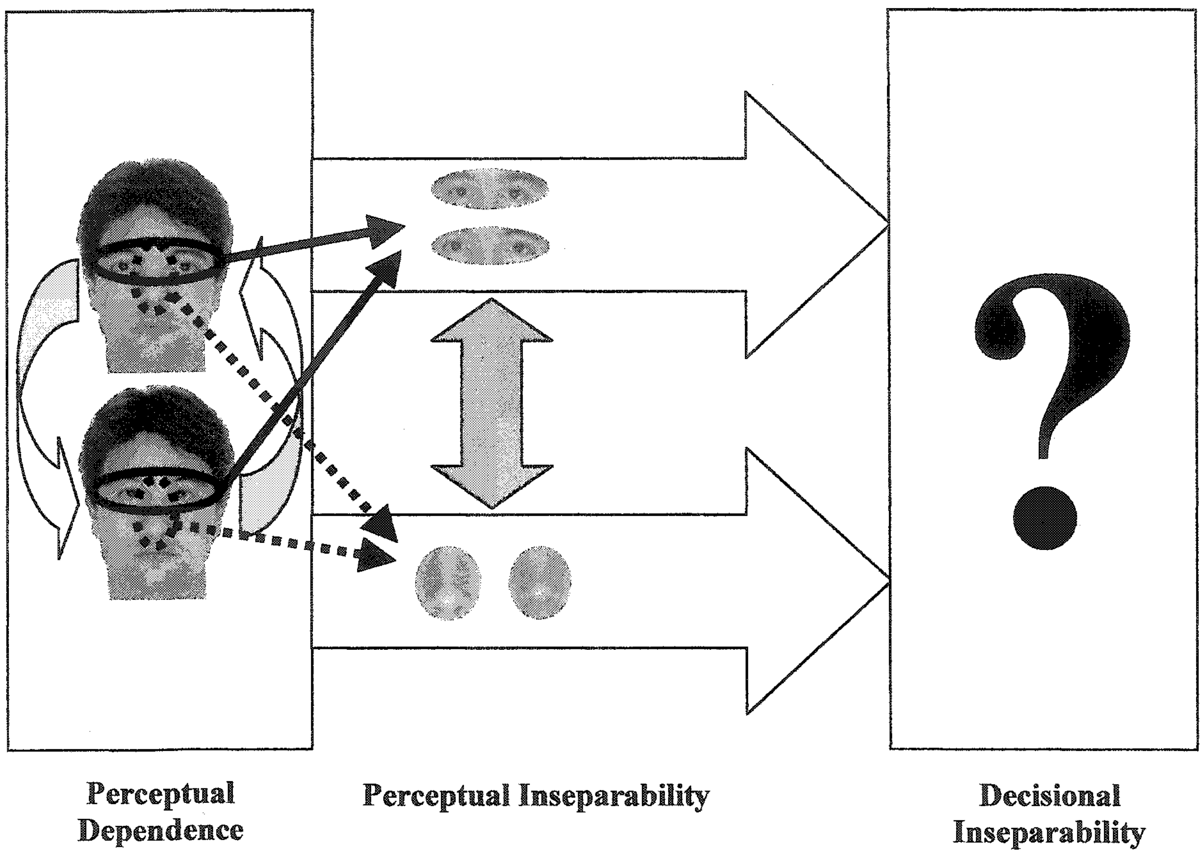

Figure 11. Two-dimensional Gaussian distribution of a hypothesized face set

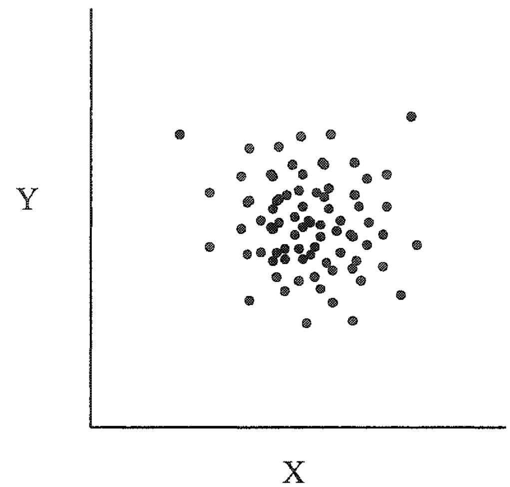

Figure 12. A hypothesized three-dimensional mental face space, where the third dimension is calculated based on the two physical dimensions (f(x,y)).