Title
Distinguishing Between Perceptual and Decisional Sources of Holism in Face Processing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Richler, Jennifer J.
Mack, Micheal L.
Gauthier, Isabel
et al.

Publication Date
2007

Peer reviewed

Distinguishing Between Perceptual and Decisional Sources of Holism in Face Processing

Jennifer J. Richler (jennifer.j.richler@vanderbilt.edu)
Psychology Department, 301 Wilson Hall
Nashville, TN 37203 USA

Michael L. Mack (michael.mack@vanderbilt.edu)
Psychology Department, 301 Wilson Hall
Nashville, TN 37203 USA

Isabel Gauthier (isabel.gauthier@vanderbilt.edu)
Psychology Department, 301 Wilson Hall
Nashville, TN 37203 USA

Thomas J. Palmeri (thomas.j.palmeri@vanderbilt.edu)
Psychology Department, 301 Wilson Hall
Nashville, TN 37203 USA

Abstract

Face recognition is thought to rely on holistic processing, where the entire face is processed as a single unitary object. The nature of this holism is often assumed to be perceptual, however recent work applying the General Recognition Theory (GRT) framework suggests that holistic processing of faces may arise from decisional factors. Using Monte Carlo simulations, we examined how known violations of GRT constructs relate to a behavioral measure of holism, the congruency effect, and found that both perceptual and decisional sources of holistic processing can give rise to significant congruency effects. We then explored whether a well-known model of face recognition, one that reproduces the congruency effect, would also reproduce the pattern of GRT results seen in human participants. Surprisingly, like humans, the model showed a consistent decisional locus of holistic processing. This was surprising because there is no explicit decisional component in the model that could give rise to this finding. This suggests that either the model contains some kind of implicit decisional component, or alternatively that there may problems with the techniques used to measure decisional versus perceptual holism from the perspective of GRT.

Keywords: holistic processing; face recognition

Introduction

A key characteristic of face perception which sets it apart from regular object perception is that faces are said to be processed holistically. That is, the features which make up a face are not treated as independent parts, but rather the eyes, nose and mouth are fused together and treated as a whole. Typically, holistic processing of faces has been attributed to holism within the perceptual representation (e.g., Young, Hellawell & Hay, 1987; Hole, 1994), a perceptual fusion of the face parts during recognition, however it is also possible that important holistic effects may arise because of a form of holism within a decisional process (Wenger & Ingvalson 2002, 2003).

The purpose of this paper is to 1) explore how one behavioral measure of holism, the *congruency effect*, relates to either perceptual or decisional loci of holistic effects as defined within General Recognition Theory (GRT; Ashby & Townsend, 1986), and 2) determine whether one current model of face recognition, one that gives rise to congruency effects similar to those seen in humans, localizes holism as either perceptual or decisional using GRT measurement tools. Results will be related to recent human data that measures holistic processing both in terms of the congruency effect and in terms of GRT constructs.

Behavioral Measure of Holism: The Congruency Effect

Holism in face recognition has been measured behaviorally using a *composite face task*. This sequential matching task uses composite faces that are created by combining the top half of one face with the bottom half of another face. Participants study one composite face and then are tested on a second composite face after a brief delay. In one version of this paradigm, participants judge whether the top or bottom of a test face is the same or different as the study face; the cued part can be the same or different, and the irrelevant part can also be the same or different. *Congruent trials* are those in which the top and bottom are both the same or both different whereas *incongruent trials* are those in which one part is the same and the other part is different. Holism is inferred by a *congruency effect*, which is defined as better performance (measured in terms of d') on congruent vs. incongruent trials. Significant congruency effects are seen with faces but not common objects. Even though participants are instructed to ignore the irrelevant part, it still affects their performance. One explanation of these results is that the irrelevant part cannot be ignored because the face is processed as a unitary perceptual whole (Farah, Wilson, Drain & Tanaka, 1998; Gauthier, Curran, Curby & Collins, 2003).

In most studies that measure holism in terms of a congruency effect, only a single response is made to either the top or the bottom of the test face on every trial (e.g., Farah et al., 1998; Gauthier et al., 2003). However, a slight variation of this task requires participants to give a same-different response to both the top and the bottom of the test face on every trial (Wenger & Ingvalson, 2002, 2003; Richler, Gauthier, Wenger, & Palmeri, submitted). This task produces similar congruency effects as the single-response version, but it also allows the data to be analyzed in terms of GRT constructs (described below) which distinguish between perceptual and decisional loci of holistic effects. For this reason, we will be referring to this version of the composite task in the remainder of the paper.

General Recognition Theory

According to classic signal detection theory, a participant's response to a stimulus reflects both a perceptual process and a decision process. On each trial, the perceptual effect of a stimulus (output of the perceptual process) can be represented as a point in perceptual space. Because of perceptual noise, that perceptual effect varies over trials, resulting in a probability distribution of percepts, which is commonly assumed to be normal. Discriminability between stimuli is reflected by the distance between perceptual distributions. Responses – and response biases – are determined by where a percept lies in relation to a decision criterion placed between the distributions.

Signal detection theory can readily be applied to performance in a sequential same-different task. But in this case, there is a percept of the first stimulus, a percept of the second stimulus, and a comparison process that computes a measure of similarity between the first and second stimulus. Consider a simple face matching task, where the participant studies one face and is tested on a second face, and simply needs to judge whether the two faces are the same or not. There would be one distribution for "same" trials and one distribution for "different" trials. A response criterion would determine whether the participant responds "same" or "different". The proportion of hits (i.e., responding "same" when the correct response is "same") and false alarms (i.e., responding "same" when the correct response is "different") are used to calculate both d', which measures disciminability between "same" and "different" trials, and the response criterion, which reflects response biases. In order to generalize this analysis to the sequential composite face paradigm that is of interest to us, we need a multidimensional analysis that can reflect the same-different status of both the top part and the bottom part.

General Recognition Theory (GRT; Ashby & Townsend, 1986) is a multidimensional generalization of classic signal detection theory. Stimuli give rise to percepts drawn from multivariate normal distributions. Responses are determined by decision boundaries separating these distributions. Within this framework, holism can be described as having a perceptual locus, by violations of *Perceptual Independence* (PI) or violations of *Perceptual Separability* (PS), or as having a decisional locus by violations of *Decisional Separability* (DS). Each of these constructs and how they can be violated are illustrated in Figure 1.

Two stimulus dimensions are perceptually independent when the perceptual effect of one dimension (or part) is statistically independent of the perceptual effect of another dimension (or part). If faces exhibit PI, then variability in the perceived sameness of the top part would be uncorrelated with variability in the perceived sameness of the bottom part. This is illustrated by the circular contours of equal likelihood in Figure 1a (left). PI is violated when there is covariance between the perceptual dimensions, resulting in the tilted ellipses seen in Figure 1a (middle). In other words, variability in the judged sameness of the top and bottom parts would be correlated with one another. Unlike some of the other violations to be discussed, PI is considered a within-stimulus effect; some intrinsic property of perceptual processing gives rise to correlated noise across the two parts of the face. Because of this, a violation of PI has been considered to be the strongest form of holism within the GRT framework (Wenger & Ingvalson, 2002).

Perceptual distributions are perceptually separable when the perception of one dimension is independent of the level of the other dimension. If faces exhibit PS, then the perceived sameness of the top part would be unaffected by whether the bottom part is the same or different. As shown in Figure 1b (left), this can be illustrated by connecting the centers of the four perceptual distributions into a rectangle. PS is violated when the perception of one part of the stimulus depends on the level of the other part. As shown in Figure 1b (middle), this can be illustrated when the connected perceptual distributions form a non-rectangular quadrilateral. If faces violate PS, there may be differences in the degree of sameness for face bottoms when the tops are the same compared to when the tops are different.

Finally, responses to each part of the stimulus are decisionally separable when the location of the bound for decisions about one dimension is not affected by the level of the other dimension. If DS applies to faces, then the boundary established for decisions about the top part is in the same location irrespective of whether the bottom part is the same or different. As shown in Figure 1c (left), this can be illustrated by linear decision bounds that are parallel to dimensional axes. Also as shown in Figure 1c (middle), when DS is violated, the location of the decision bound for one part depends on the other part. If faces violate DS, the location of the criterion to say whether bottoms are the same or different may vary based on whether the top is the same or different.

Simulations Relating Violations of GRT Constructs With the Congruency Effect

Recall that the standard measure of holistic processing, the congruency effect, is often attributed to a perceptual form of holism. But within GRT, holism could be perceptual (violations of PI or PS) or decisional (violations of DS). Are congruency effects sensitive to these violations? If the congruency effect can *only* arise due to violations of PI or PS, or *only* arise due to violations of DS, GRT constructs tell us little more than what is already measured by congruency effects, because the congruency effect on its own would be exclusively linked to a perceptual or decisional locus of holistic processing. However, if congruency effects can be produced by violations of *any* of these constructs, then analyzing data using GRT constructs could prove to be a more powerful analytic tool for understanding the nature of holistic processing of faces.

To answer these questions, we performed a series of Monte Carlo simulations where we systematically violated PI, PS, and DS and examined the congruency effect that emerged in each simulation. The middle column of Figure 1 illustrates how each of the GRT constructs was manipulated. We assumed a multivariate distribution for each of the combinations of same or different top with same or different bottom of the test face. PI was systematically manipulated by varying the correlation (rho) between the two dimensions in the covariance matrix (Ashby, 1992); when there is a zero correlation, there is no violation of PI. "Squareness" of the configuration of perceptual dimensions (theta) was varied to systematically manipulate PS; when theta is set to 45 degrees the arrangement of the distributions is a square and there is no violation of PS. We systematically manipulated DS by changing the orthogonality of the decision bounds to the dimensional axes (phi); when phi is set to 0, the decision bounds are orthogonal to the coordinate axes and there is no violation of DS. We also manipulated the variance (var) along the two perceptual dimensions, but this parameter simply acts as a scaling term and has no qualitative effect. Each GRT construct was investigated independently; for example, if we were examining a violation of PI with rho set to .5, we assumed no violation of PS (theta = 45) and no violation of DS (phi = 0).

For each set of parameters, we ran a total of 4000 simulated trials. On each trial, we randomly selected one of the four multivariate normal distributions, and then randomly drew a sample from that distribution. The response for each trial was determined based on where the sample was located with respect to the decision boundaries. Responses could then be characterized as hits (correctly responding "same") or false alarms (responding "same" when the correct response is "different"). The proportion of

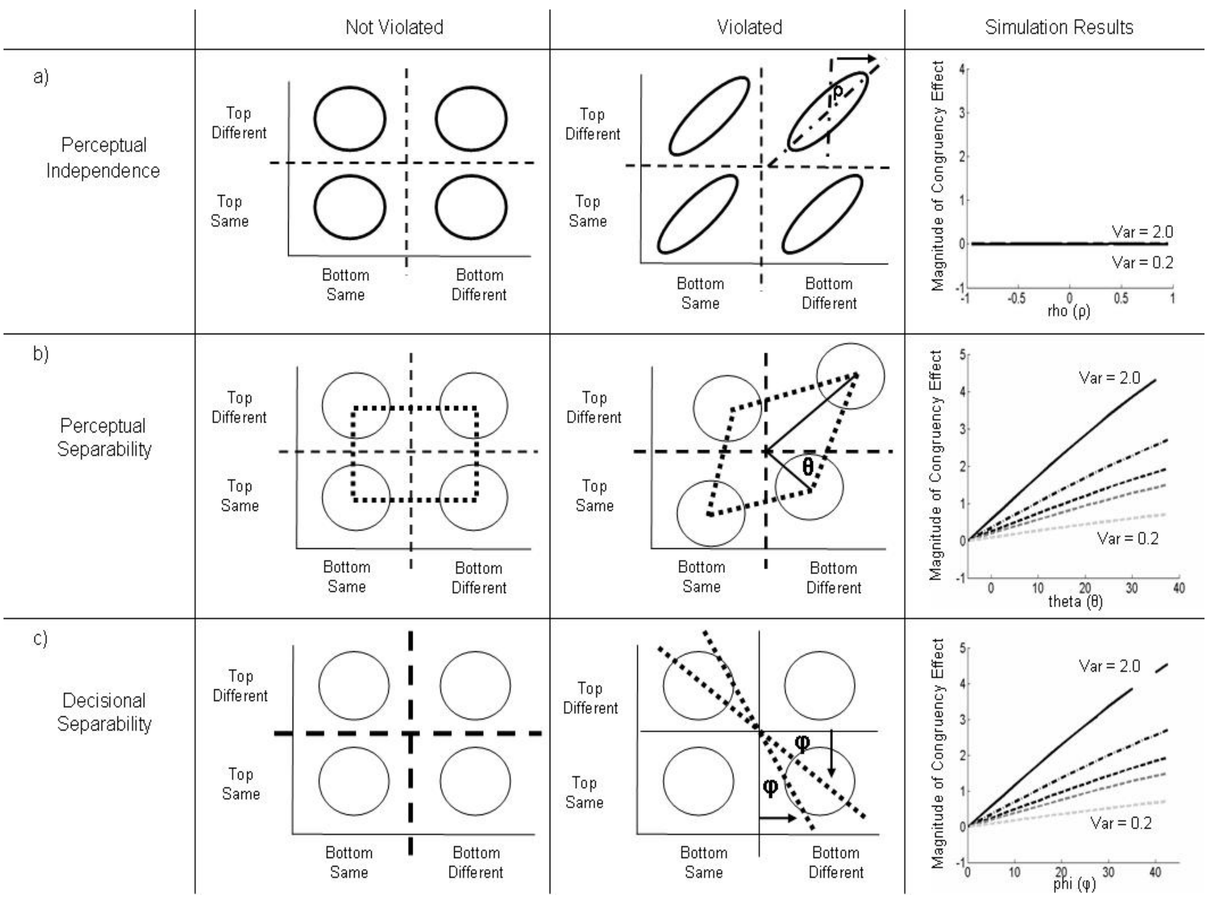

Figure 1. (a) PI (left), a violation of PI and how PI was manipulated in the simulations (middle) and results of the simulations for PI (right). (b) PS (left), a violation of PS and how PS was manipulated in the simulations (middle) and the results of the simulations of PS (right). (c) DS (left), a violation of DS and how DS was manipulated in the simulations (middle) and the results of the simulations (right).

hits and false alarms were used to calculate d’ for congruent trials (both parts “same” or both parts “different”) and incongruent trials (one part “same” and one part “different”). The magnitude of the congruency effect is the difference between d’ for congruent and incongruent trials.

The right column of Figure 1 summarizes the results of the simulations of violations of each of the three GRT constructs. Each graph plots the magnitude of the congruency effect as a function of the value of the systematically manipulated parameter for that simulated violation. Each line represents a different value of the dimensional variance (var).

As shown by the flat line in the top graph of Figure 1, no violation of PI produces a congruency effect. This is particularly surprising given that a violation of PI is considered to be the strongest form of holism. If holistic processing of faces occurs due to violations of PI, this cannot be detected by measuring the congruency effect.

Violations of PS and DS, on the other hand, both produce congruency effects, as shown by the increasing lines in the middle and bottom graphs in Figure 1. When PS is violated, the magnitude of the congruency effect increases as the configuration becomes less square-like (i.e., as theta increases). When DS is violated, the magnitude of the congruency effect increases as the decision bounds become less orthogonal to the dimensional axes (i.e., as phi increases in magnitude).

Violations of PS indicate perceptual holism and violations of DS indicate a decisional locus of holism. However, these simulations show that a congruency effect can arise due to either violations of PS and DS. Therefore, the congruency effect on its own cannot distinguish between perceptual and decisional sources of holism as defined by GRT.

Behavioral Data: GRT

Conclusions about the GRT constructs can be derived from human data using an analysis technique called Multidimensional Signal Detection Analysis (MSDA; Kadlec & Townsend, 1992; Kadlec, 1995), which draws conclusions about violations of PI, PS and DS based on a number of measures, including what are known as *marginal* and *conditional* d’ and c values. The values are calculated from the complete confusion matrix of data obtained when participants make same-different judgments about both the top and the bottom half of the face. We very briefly describe some of the main components of MSDA, but the reader is referred to Kadlec & Townsend (1992) for complete details of the theory underlying the statistical tests involved in MSDA, and to Kadlec (1995) for complete details of the MSDA program we used to perform these analyses.

Marginal tests compare d’ or c values between each level of one dimension collapsed across both levels of the other dimension. For example, for a composite face task, one marginal test would compare d’ for same vs. different tops collapsed across same and different bottoms. PS is violated when there is a difference in d’ for one part based on whether the other part is the same or different. DS is violated if there is a difference in c for one part based on whether the other part is the same or different.

Conditional analyses compare d’ and c for each dimension at a given level of the other dimension, based on whether the response to the other dimension was correct or incorrect. For example, in a composite face task, one conditional test compares the criterion for a bottom decision when the top is the same and the response is correct versus the criterion for a bottom decision when the top is same and the response was incorrect. Whether the response is correct or incorrect depends on the location of the decision bounds as well as the location and shape of the perceptual distribution. Therefore, if the conditional analyses reveal violations of DS (as shown by differences in conditional c values), one cannot make conclusions about PI, because differences in conditional d’ may be due to violations of DS, rather than violations of PI (Ashby & Townsend, 1986).

Recent behavioral work that examined holistic processing for faces with respect to the GRT constructs using the MSDA program showed that DS was consistently violated in a composite face task, while PS was inconsistently violated, and PI was rarely violated (Wenger & Ingvalson, 2002, 2003; Richler et al., submitted). Moreover, it has also been shown that changes in the congruency effect due to stimulus manipulations are linked to changes in marginal c, but not marginal d’, values (Richler et al., submitted). In other words, although the Monte Carlo simulations described earlier show that violations of either PS or DS *could* give rise to a congruency effect, the analysis of the human data using MSDA suggest that it is violations of DS that play a dominant role in producing this effect in a composite face task.

Measuring Holism in one Model of Face Recognition

Because the Monte Carlo simulations showed that congruency effects can arise from violations of either PS or DS, face recognition models that account for congruency effects need not produce the same pattern of GRT violations as human observers. Thus, the GRT framework presents an opportunity to further evaluate existing face recognition models beyond the congruency effect.

In this paper, we examined one well-known model of face recognition (Cottrell, Branson, & Calder, 2002; Dailey & Cottrell, 1998). Our intent was not to particularly criticize this model; this model has accounted for a far wider range of important object recognition, face perception, and expertise phenomena than just about any extant model in the field. Instead, our intent was to illustrate that models of face recognition must also account for regularities in the human data that are revealed using MSDA; new work will examine other models of face recognition. To be clear, GRT may not be the correct underlying model of face recognition; in fact, GRT is a framework and by itself cannot be a complete model of face recognition or any other kind of object recognition. But the measures afforded by MSDA and GRT can provide more stringent measures of holistic processing in humans and models.

Our simulations closely follow the simulations presented in Cottrell et al. (2002) and Nguyen and Cottrell (2005), so we only briefly summarize them here. As shown in Figure 2, the model by Cottrell and colleagues first preprocesses face images through a bank of two-dimensional Gabor filters of different scales and orientations, the Gabor-filtered image is then represented by a vector of principal components (defined by a PCA from a different collection of faces), and then is classified by a two-layer neural network trained using back propagation.

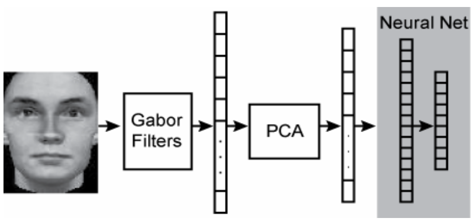

Figure 2: Model schematic adapted from Dailey, Cottrell, Padgett, & Adolphs (2002).

The training procedure was carried out in the following manner: the training sets consisted of twenty different face composites with five different exemplars of each face (created by jittering the composites by +/- 2 pixels in one or both of image dimensions). The training set was first used to generate the PCA layer representation. The model was next trained for fine-level discrimination by learning to identify the face composites with a unique name. This involved learning the weights between the PCA layer and the hidden layer and the weights between the hidden layer and the name layer via back-propagation.

To compare the model to the human data, the model had to simulate all key elements of the composite face task with same or different responses to both the top and bottom halves. Doing this involved just a simple elaboration of the approach Cottrell et al. (2002) used in their initial simulations of the task that demonstrated the model's account of the congruency effect. Each trial consisted of passing a study face through the model and saving in memory the representation of this face at the hidden layer of the model's neural network (this is akin to some form of visual working memory). As with the task used with human observers, the test face could have both parts the same as the study face, one part the same and one part different, or both parts different. Following Cottrell et al. (2002), attention to the cued part (top or bottom) was simulated by attenuating the Gabor layer representation of the to-be-ignored face half by a factor of 0.125. The correlation between the hidden layer representation of the study phase and the hidden layer representation of the test faces was used to generate a same or different response (to that half) by comparing that correlation with a criterion. Correlations higher than the criterion yield "same" responses, those lower than the criterion yield "different" responses. The process was repeated for the other half in order to generate the other same/different response. The testing stage consisted of one hundred trials each with a different study face.

Ten separate sets of training and testing faces were created and passed through the model with different random initial weights in the neural network of the model for each set. The correlation data for each model repetition was then used for calculating probability distributions for the different responses allowing for MSDA.

Comparing the Model with Human Observers

Like human observers, the model produces a significant congruency effect, whereby d' is higher on congruent trials (when both parts are the same or different) than incongruent trials (when one part is the same as the other part is different). Also similar to human observers, when the top and bottom parts of the test image are misaligned, this congruency effect diminishes (see top Figure 3). This is exactly what Cottrell et al. (2002) reported. What about the MSDA analyses?

The qualitative marginal analyses revealed consistent violations of DS and very few violations of PS in the ten simulations of the model. This matches human data, where DS is always violated and PS is violated in some cases but not in others (Wenger & Ingvalson, 2002, 2003; Richler et al., submitted).

Although the qualitative results are informative, a more rigorous comparison of the model and human observers is available when we examine the quantitative measures used to make these qualitative statements. Recall that in the marginal analyses a violation of PS or DS occurs when there is a significant difference in d' or c, respectively, based on whether the other part is the same vs. different. Richler et al. (submitted) found that stimulus manipulations that changed the congruency effect for human observers also changed marginal c values, but not marginal d' values. According to MSDA, this suggests violations of DS as a source of holistic processing. We performed the same comparisons with the model simulations. Specifically, we compared the change in the congruency effect produced by the model under the different levels of stimulus alignment with the marginal d' and c values from the MSDA output. The averages of these values are plotted in Figure 3.

The congruency effects produced by the model for the

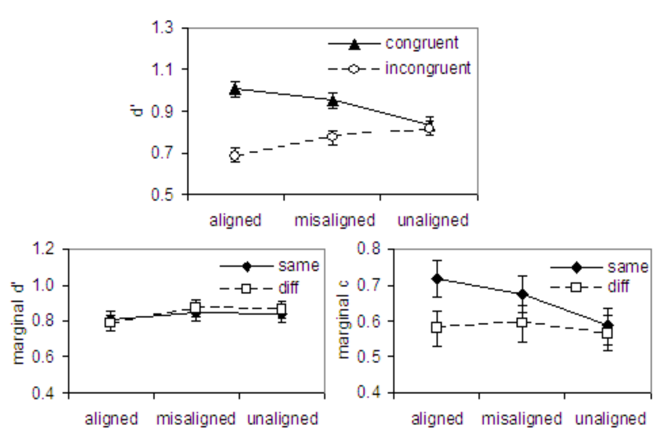

Figure 3. Congruency effect (top), marginal d' values (bottom left), and marginal c values (bottom right) for the model. Error bars show 95% confidence intervals.

three levels of alignment are shown in the top graph of Figure 3. Marginal d' values, which reveal violations of PS for the model are plotted in the bottom left, and marginal c values, which reveal violations of DS, are plotted in the bottom right. The model was able to produce the same pattern of results reported by Richler et al. with human observers. That is, marginal d' values were unaffected by misalignment. However, marginal c values were affected by misalignment such that differences in marginal c values based on the same-different status of the other part decreased with misalignment. The correspondence between the decrease in the congruency effect and the decrease in the differences between marginal c values could certainly suggests that violations of DS play a dominant role in the holistic processing effects seen in the composite task for this model. Critically, these results are consistent with findings from human observers.

Conclusions

The first goal of this paper was to relate a behavioral measure of holism used in a composite face task with measures of holism based on GRT that distinguish between perceptual and decisional loci of holistic effects. The results of the Monte Carlo simulations showed that while violations of PI – the "strongest" form of holism – do not lead to any congruency effect, both violations of PS (perceptual holism) and violations of DS (decisional holism) lead to significant congruency effects. The congruency effect is a rather coarse measure of holism in that it cannot distinguish between perceptual and decisional sources of holistic effects.

The second goal of this paper was to examine how one well-known face recognition model, one that accounts well for the congruency effect, would compare with human observers in terms of the measures of holism provided by MSDA. As is seen with human data, the model showed consistent violations of DS. This was very surprising as there is no explicit decisional component in the model that could obviously lead to this finding. This suggests that either the model contains some kind of implicit decisional component, or, alternatively, this indicates that the values derived using MSDA do not necessarily reflect the decisional sources of holism defined by GRT. We are currently exploring both possibilities.

We conclude that 1) behavioral work that seeks to explain holistic processing of faces needs to use measures that are more sensitive to the distinction between perceptual and decisional sources of holism, such as MSDA, and 2) a more thorough investigation of MSDA and how it relates to GRT is required.

Acknowledgments

This work was supported by NSF SBE-0542013, NSF HSD-DHBS05, and a grant from the James S. McDonnell Foundation. Thanks to Gary Cottrell and the GURU lab for providing the software implementation of their model for use in this research.

References

Ashby, F.G. (1992). Multivariate Probability Distributions. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Erlbaum.

Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154-179.

Cottrell, G.W., Branson, K., & Calder, A.J. (2002). Do expression and identity need separate representations? *In Proceedings of the 24th Annual Cognitive Science Conference*, Fairfax, Virginia. Mahwah: Lawrence Erlbaum.

Dailey, M.N., & Cottrell, G.W. (1998). Task and spatial frequency effects on face specialization. *Advances in Neural Information Processing Systems 10*, MIT Press, Cambridge, MA.

Dailey, M.N., Cottrell, G.W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14(8)*, 1158-1173.

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

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

Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23,* 65-74.

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., & Townsend, J.T. (1992). Signal detection analysis of dimensional interactions. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Nguyen, N., & Cottrell, G.W. (2005). Owls and wading birds: Generalization gradients in expertise. In *Proceedings of the 27th Annual Cognitive Science Conference*, La Stresa, Italy. Mahwah: Lawrence Erlbaum.

Richler, J.J., Gauthier, I., Wenger, M.J., & Palmeri, T.J. (submitted). Holistic processing of faces: Bridging paradigms.

Wenger, M.J., & Ingvalson, E.M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 872-892.

Wenger, M.J., & Ingvalson, E.M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 1106-1118.

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