# Psychological Science

The Perception of a Face Is No More Than the Sum of Its Parts

Jason M. Gold, Patrick J. Mundy and Bosco S. Tjan Psychological Science published online 5 March 2012 DOI: 10.1177/0956797611427407

The online version of this article can be found at: http://pss.sagepub.com/content/early/2012/03/02/0956797611427407

Published by: SAGE http://www.sagepublications.com

On behalf of: On behalf of:

Association for Psychological Science

Additional services and information for *Psychological Science* can be found at:

Email Alerts: http://pss.sagepub.com/cgi/alerts

Subscriptions: http://pss.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> OnlineFirst Version of Record - Mar 5, 2012 What is This?

# The Perception of a Face Is No More Than the Sum of Its Parts

# Jason M. Gold<sup>1</sup>, Patrick J. Mundy<sup>1</sup>, and Bosco S. Tjan<sup>2</sup>

<sup>1</sup>Indiana University and <sup>2</sup>University of Southern California

## Abstract

Psychological Science XX(X) 1–8 © The Author(s) 2012 Reprints and permission: sagepub.com/journalsPermissions.nav DOI: 10.1177/0956797611427407 http://pss.sagepub.com

CHOLOGICAL SCIENCE



When you see a person's face, how do you go about combining his or her facial features to make a decision about who that person is? Most current theories of face perception assert that the ability to recognize a human face is not simply the result of an independent analysis of individual features, but instead involves a holistic coding of the relationships among features. This coding is thought to enhance people's ability to recognize a face beyond what would be expected if each feature were shown in isolation. In the study reported here, we explicitly tested this idea by comparing human performance on facial-feature integration with that of an optimal Bayesian integrator. Contrary to the predictions of most current notions of face perception, our findings showed that human observers integrate facial features in a manner that is no better than would be predicted by their ability to use each individual feature when shown in isolation. That is, a face is perceived no better than the sum of its individual parts.

#### **Keywords**

visual perception, face perception, perception, vision

Received 6/27/11; Revision accepted 9/30/11

One aspect of face perception that has received a considerable amount of attention in recent years is how the different features of a face (e.g., eyes, nose, mouth) are processed and represented by the visual system. One possibility is that each facial feature is analyzed independently, and recognition involves simply integrating all of the different elements of a face (i.e., a piecemeal analysis; Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Grand, & Mondloch, 2002). Another possibility is that, similar to the Gestalt notion that the whole is more than the sum of its parts, a face is analyzed as a single unified entity, and the spatial relationships among the features are encoded as part of the representation (i.e., a holistic analysis). This process can allow an observer to make better use of information than if each of the individual features is represented in isolation (Farah et al., 1998; Maurer et al., 2002; Richler, Cheung, & Gauthier, 2011).

Several lines of evidence are consistent with the idea that faces are processed in a holistic rather than a piecemeal manner. Much of this evidence draws on a phenomenon called the *face-inversion effect*—the finding that, unlike most other objects, faces tend to be much more difficult to identify when they are inverted than when they are upright (Maurer et al., 2002; Richler, Mack, Palmeri, & Gauthier, 2011; Valentine, 1988; Yin, 1969). This effect is typically accounted for by positing that upright faces are processed in a holistic manner, whereas the elements of inverted faces are processed in a piecemeal manner. As a result, the extra information that is encoded for upright faces allows an observer to identify an upright face more quickly and accurately than an inverted face. Other experiments (Tanaka & Farah, 1993) have shown that observers are more accurate at identifying facial features (e.g., a nose) within the context of a normal face than either in isolation or in the context of a face whose features have been spatially scrambled. However, some researchers have suggested that phenomena such as the face-inversion effect simply reflect a quantitative shift in the efficiency of information use rather than a qualitative shift in recognition strategy (Sekuler, Gaspar, Gold, & Bennett, 2004).

One problem with many such studies is that the concept of holism is often expressed in purely descriptive terms, so it can be difficult to make clear quantitative predictions about the results of experiments. In the study reported here, we took an alternative approach and compared human performance on

**Corresponding Author:** 

Jason M. Gold, Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th St., Bloomington, IN 47405 E-mail: jgold@indiana.edu

facial-feature integration with that of an optimal Bayesian integrator (Ernst & Banks, 2002; Knill & Saunders, 2003; Stimuli Landy & Kojima, 2001; Nandy & Tjan, 2008). The advantage of this approach is that it offers a clearly formalized frame-

work and has been successfully applied to several different domains other than face perception (Ernst & Banks, 2002; Graham, Robson, & Nachmias, 1978; Hillis, Ernst, Banks, & Landy, 2002; Knill & Saunders, 2003; Landy & Kojima, 2001; Nandy & Tjan, 2008; Pelli, Farell, & Moore, 2003; Sorkin, Hays, & West, 2001).

An optimal Bayesian integrator is a theoretical observer that combines information from different sources in a statistically optimal manner. More specifically, given a particular observer's performance with each of a series of individual sources of information, the optimal Bayesian integrator yields the performance that would be expected if all of the individual sources of information were integrated in a statistically optimal fashion when presented in combination. Performance with the combination that is worse than predicted by the optimal Bayesian integrator implies that information is lost through the integration process (suboptimal integration). Performance that is better than predicted by the optimal Bayesian integrator (superoptimal integration) implies that information is lost when each individual source is shown in isolation, in the absence of the integration process. Superoptimal integration therefore provides a distinct behavioral definition for holistic processing, in the sense that the whole is more than what would be predicted from the sum of the parts.

This general approach has been highly successful in elucidating how information is combined across multiple depth and spatial cues (Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Landy & Kojima, 2001); different modalities, such as vision and touch (Ernst & Banks, 2002); letters within words (Pelli et al., 2003); spatial frequencies (Graham et al., 1978; Nandy & Tjan, 2008); and even members of a group making collective decisions (Sorkin et al., 2001). In the context of how information is combined across facial features, the optimal Bayesian integrator makes a strong prediction: If observers use a holistic face representation, in which an intact face allows the observer to make better use of information than when the parts are presented in isolation, we would predict that an observer would exhibit superoptimal feature-integration performance. In contrast, if the observer has the same ability to use information whether it is from a whole face or from isolated face parts, we would not expect the observer's feature-integration performance to exceed that of an optimal Bayesian integrator.

To test these predictions, we conducted a series of three experiments, in which observers were asked to identify the facial features (mouth, nose, left eye, right eye, shown either in isolation or in combination) of a set of six faces. By comparing performance with the isolated features to that with the combination, we were able to explicitly test whether observers integrated information across features in a fashion that exceeded the predictions of an optimal integrator (i.e., superoptimal integration).

# **General Method**

In each experiment, observers performed a one-of-six identification task. In this task, one of six possible images was presented against a gray background on a computer screen. Observers were then shown all six images simultaneously, and they had to decide which of the six had just been shown separately.

Each observer participated in five different conditions (Fig. 1). In the combined condition, we extracted facial features from  $2.5^{\circ} \times 2.5^{\circ}$  gray-scale photographs of three male and three female faces used in previous experiments on face recognition (Gold, Bennett, & Sekuler, 1999a, 1999b). Four small Gaussian windows of constant size ( $\sigma = 0.1^{\circ}$ ) were applied at four fixed locations on each face: the left eye, the right eye, the nose, and the mouth. These windows revealed the parts of the faces that were directly below the centers of the windows, with a smooth falloff in contrast away from the centers to the background. In the other four conditions, only one of the four features from each of the combined images was shown, but the position in which it appeared was exactly the same as in the combined condition. All five conditions (right eye only, left eye only, nose only, mouth only, all features combined) were randomly intermixed across trials throughout testing.

# Design and procedure

On each trial, an image was randomly chosen and presented on a computer screen for 500 ms in low-contrast Gaussian pixel noise ( $\sigma = 0.001$ ), after which a response screen appeared containing only the six images from that condition (e.g., if only a nose was shown on a given trial, the selection window contained the six possible nose-only images). Observers had to choose which of the six images matched the one that had just been presented. The visual noise was perceptually inconsequential; it was included for the technical reason of imposing a limit on ideal-observer performance.

The contrast of the images was varied across trials within a given session according to five interleaved adaptive staircases. Each session consisted of 600 trials (120 trials per condition). Each observer completed five sessions over the course of 5 days of testing. We discarded the first two sessions, as significant learning has been shown to take place during these initial sessions in similar face-perception tasks (Gold et al., 1999b; Gold, Sekuler, & Bennett, 2004). We fitted psychometric functions to the staircase data from the last three sessions to obtain 50%-correct contrast thresholds in each condition (chance performance was 17% correct).

# The optimal Bayesian integrator and the integration index

Several different approaches have been used to derive the predictions of an optimal Bayesian integrator. Although all of these approaches are based on the same principles, the mathematical formalization that is required to arrive at the optimal



**Fig. 1.** Examples of the stimuli used in Experiments I through 3. Four features were isolated from six different face images (each column shows features from a single face). In the combined condition, all four features were shown. In the other four conditions, only one feature from a face was shown. In Experiment I, features were presented upright against a gray background. In Experiment 2, all features were presented upright inside of a background face that was created by averaging the whole faces from which the features were taken and leaving four holes for the features; the examples in the bottom row show the features from the combined condition presented against the average-face background. In Experiment 3, the five conditions were the same, but the features were inverted; half of the observers saw features against a gray background, and the other half saw features against the average-face background.

prediction varies according to the specifics of the task and the stimuli, as well as how performance is quantified. In the case of the task used here, it has been shown that the performance of the optimal Bayesian integrator is such that the squared sensitivity in the combined condition is equal to the sum of the squared sensitivities in each of the single-feature conditions. We can therefore define an integration index,  $\Phi$ , as follows:

$$\Phi = \frac{S_{\text{combined}}^2}{S_{\text{left eye}}^2 + S_{\text{right eye}}^2 + S_{\text{mouth}}^2 + S_{\text{nose}}^2} , \qquad (1)$$

where *S* is the reciprocal of the observer's contrast threshold. The contrast of a face feature is defined as the contrast of the intact face from which the feature is extracted (its *nominal contrast*). That is, the combined face image is first set to the specified level of contrast, and then individual features are extracted from this image according to the specific condition. With this definition of contrast, the integration index,  $\Phi$ , is equal to 1 if observers are integrating information optimally,

less than 1 if they are integrating information suboptimally, and greater than 1 if they are integrating information superoptimally (Nandy & Tjan, 2008).

# Experiment 1: Upright Faces Method

In our first experiment, 5 observers performed our identification task in the five conditions described in the General Method, displayed in the familiar upright orientation (Fig. 1).

#### Results

The results of Experiment 1 are shown in Figures 2a and 2b. Figure 2a plots the sensitivities for the 5 observers in all conditions, along with the sensitivity of a statistically optimal pattern classifier (also known as an *ideal observer*) in the same conditions (Gold et al., 1999a; Tjan, Braje, Legge, & Kersten, 1995). The performance of the ideal observer is constrained



**Fig. 2.** Results of Experiment 1 (a, b) and Experiment 2 (c, d). The graphs in the left column show contrast sensitivity as a function of condition for 5 human observers and an ideal observer. The graphs in the right column show the integration index as a function of observer, along with the mean across observers. Also shown is the integration index predicted by a suboptimal best-feature model observer (see the text). In Experiment 1, facial features were shown in an upright position against a gray background. In Experiment 2, facial features were shown in an upright position against a gray background. In Experiment 2, facial features were shown in an upright position against a gray background face image. Error bars on all individual sensitivities and indices were obtained through bootstrap simulations (Efron & Tibshirani, 1993) and represent  $\pm 1$  SD in the case of sensitivity and  $\pm 1$  SD in the case of the integration index. Error bars for the mean integration indices show  $\pm 1$  SEM. The optimal index is 1, which is highlighted by the dashed horizontal line.

only by the intrinsic difficulty of the task and therefore represents a strict upper boundary on performance in each condition at the given noise level. These data show that the pattern of sensitivities across conditions for most human observers was quite similar to that of the ideal observer.

Figure 2b plots the corresponding integration index for each observer, the mean index across observers, and the corresponding integration index for the ideal observer, derived from the sensitivities shown in Figure 2a. As expected, the integration index for the ideal observer was equal to 1, because it always used all of the available information in an optimal fashion. The graph also shows the integration index predicted by a suboptimal *best-feature* model observer. The best-feature observer is constrained to use only the single feature that has the highest sensitivity in isolation, and hence requires the least amount of contrast, when performing in the combined condition. Although there was considerable variability across observers, the mean index for the human observers was 0.82, which did not significantly differ from 1, t(4) = -1.36, p =.25.<sup>1</sup> On the basis of these results, we can conclude that, on average, our observers integrated information across facial features in a fashion that was very close to and statistically indistinguishable from optimal. Surprisingly, observers did not appear to derive any additional benefit beyond that of having four features instead of one in the combined condition.

# Experiment 2: The Effect of an Added Background Face Image Method

Could it be that the piecemeal appearance of our stimuli compelled observers to use a strategy that was also piecemeal in nature? We tested this possibility in Experiment 2 by generating a background face image that was an average of the six original face images from which the features were drawn, minus the locations where the Gaussian windows appeared (Fig. 1, bottom row). This background face image was added to all of the images used in Experiment 1. Adding a constant background image introduces no additional information to the task, and therefore has no effect on the sensitivity of the ideal observer. However, the inclusion of the background face image provides context for the facial features, and makes the images appear much more facelike. We tested 5 new observers with these stimuli in the same conditions as in Experiment 1.

# Results

The results of Experiment 2 (Figs. 2c and 2d) were very similar to those obtained without the background face image in Experiment 1. Most notably, the mean index with the background face image was nearly identical to the mean index from the first experiment (0.84) and was marginally less than 1, t(4) = -2.45, p = .07. A between-subjects analysis of

variance (ANOVA) applied to the data from Experiments 1 and 2 revealed no significant effect of the background face image on the integration index, F(1, 8) = 0.24, p = .64. As such, we combined the data from Experiments 1 and 2 to compute the overall mean index (0.83), which was marginally less than 1, t(9) = -2.17, p = .06. Thus, we found no evidence of superoptimal integration of facial features in our first two experiments. If anything, observers appeared to be slightly suboptimal in their ability to combine features across a face.

# Experiment 3: The Effect of Inversion Method

In our third experiment, we investigated the impact that inverting facial features would have on integration efficiency. The negative effects of inversion in face-perception tasks are typically taken as strong evidence for holistic face processing (Farah, Tanaka, & Drain, 1995; Maurer et al., 2002; Valentine, 1988; Yin, 1969). We created a new set of features that were identical to our previously generated stimuli (both with and without the average-face background image), except that each stimulus was flipped upside down. We tested 10 new observers in the same conditions as in Experiments 1 and 2. Observers were divided evenly into two groups, with one group viewing only the inverted features without the added background face image used in Experiment 1, and the other viewing only the inverted features with the added background face image (also inverted) used in Experiment 2.

## Results

Figure 3 shows the results of Experiment 3. The mean index for the observers was 0.51 for inverted features with no background faces (Fig. 3b) and 0.61 for inverted features with background faces (Fig. 3d). Both indices were significantly less than 1, inverted features only: t(3) = -5.08, p < .02; inverted features with background faces: t(5) = -3.53, p < .02). A 2 (face orientation: upright, inverted) × 2 (background face image: present, absent) between-subjects ANOVA showed a significant main effect of orientation, F(1, 16) = 5.13, p < .05, and no main effect of background face image, F(1, 16) = 0.30, p = .59. There was no interaction between orientation and background face image, F(1, 16) = 0.30, p = .59.

We also compared the results of Experiment 3 with the results of the previous two experiments. The significant effect of face orientation on integration index was in stark contrast with the lack of any main effect of orientation on observers' sensitivities for individual facial features presented alone, as revealed by a 2 (face orientation) × 2 (background face image) ANOVA on the mean log sensitivities for isolated facial features, F(1, 16) = 0.033, p = .86. There was also no effect of background face image, F(1, 16) = 0.21, p = .65, but there was a significant interaction between face orientation and background face image, F(1, 16) = 11.62, p < .01, with lower



**Fig. 3.** Results of Experiment 3 for inverted facial features presented against a gray background (a, b) and against a background face image (c. d). The graphs in the left column show contrast sensitivity as a function of condition for human observers and the ideal observer. The graphs in the right column show the integration index as a function of observer, along with the mean across observers. Also shown is the integration index predicted by a suboptimal best-feature model observer (see the text). Error bars on all individual sensitivities and indices were obtained through bootstrap simulations (Efron & Tibshirani, 1993) and represent  $\pm 1$  SD in the case of sensitivity and  $\pm 1$  SD in the case of the integration index. Error bars for the mean integration indices show  $\pm 1$  SEM. The optimal index is 1, which is highlighted by the dashed horizontal line.



**Fig. 4.** Mean integration index across all three experiments as a function of stimulus type. Also shown is the integration index predicted by a suboptimal best-feature model observer (see the text). Facial features were presented either upright or inverted and against either a gray background or a background face image. Error bars show  $\pm I$  SEM. The optimal index is I, which is highlighted by the dashed horizontal line.

sensitivities for the conditions with either upright features in the absence of any background face or inverted features in the presence of an inverted background face. Figure 4 shows the mean index across observers in Experiments 1 through 3 for both upright and inverted faces with and without the presence of the background face image.

# Discussion

Three conclusions can be drawn from our experiments. First, our results are inconsistent with the idea that the individual parts of a face are processed in a qualitatively different manner when they are shown in isolation as opposed to shown in combination. If this idea were correct, we would have expected to see clear evidence for superoptimal integration efficiency, which literally means that performance with the whole would have been better than the sum of performance with the individual parts. Instead, we found that observers integrated information across an upright set of face features in a fashion that was predicted (or perhaps slightly underpredicted) by their performance with the parts presented in isolation.

Second, the face-inversion effect also does not seem to be due to a qualitative shift in face processing. We are able to draw this conclusion because we found no evidence of superoptimal integration with upright features. As such, the lower integration efficiency observed with inverted features than with upright features implies that inverted features are simply integrated in a less efficient manner than upright features are. Although at first it may seem that this result implies that the spatial arrangement of features plays no role in the perception of a face, the fact that integration efficiency was reduced for inverted features actually provides strong evidence that the configuration of features does matter. But it forces a reconsideration of how to interpret the impact of the spatial arrangement of features on face perception. Our results indicate that the configuration of features in a face affect how much information from each facial feature one can utilize to identify a face. However, regardless of the configuration of the features (orientation in our study), we found that the information observers utilized was up to and no more than the information they were able to make use of when each feature was presented individually. That is, the configuration of facial features has a quantitative impact on integration efficiency rather than a qualitative effect on processing strategy. This result is consistent with some previous findings that the weights observers assign to different parts of a face are more similar to those used by an ideal observer when faces are upright as opposed to inverted (Sekuler et al., 2004). However, it still remains an open question as to why the integration process is disrupted by inversion.

Finally, it is worth pointing out that there is some evidence that the degree of familiarity an observer has with a face may play a role in how a face is processed (e.g., Dubois et al., 1999; Ellis, Shepherd, & Davies, 1979). Although each observer who participated in our experiments underwent two extensive sessions of initial familiarization with the faces, it remains a question to be determined by future research whether our results extend to highly familiar faces. Nevertheless, the results of our experiments show that for newly familiar faces in the common upright orientation, an observer's ability to use information is nearly the same for individual facial features as it is for an intact face. In other words, the perception of a face is no more than the sum of its parts.

#### Acknowledgments

We would like to thank Megan Shotts for help in data collection. We would also like to thank Mike Burton and one anonymous reviewer for helpful comments.

### **Declaration of Conflicting Interests**

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

#### Funding

This research was supported by National Institutes of Health Grants R01-EY019265 (to J. M. G.) and R01-EY017707 (to B. S. T.).

#### Note

1. All statistical tests were conducted with log index or log sensitivity, because measurement noise in contrast thresholds (i.e., 1/sensitivity) typically follows a log normal distribution. We also conducted all tests in linear units, and in every case, the results were qualitatively the same.

#### References

- Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *NeuroImage*, 9, 278–289.
- Efron, B., & Tibshirani, R. (1993). An introduction to the bootstrap. New York, NY: Chapman & Hall.
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*, 431–439.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, 21, 628–634.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105, 482–498.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39, 3537–3560.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature*, 402, 176–178.
- Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28, 167–207.
- Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, 18, 815–825.
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298, 1627–1630.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12), Article 1. Retrieved from http://www .journalofvision.org/content/4/12/1

- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 18*, 2307–2320.
- Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6, 255–260.
- Nandy, A. S., & Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *Journal of Vision*, 8(13), Article 3. Retrieved from http:// www.journalofvision.org/content/8/13/3.short
- Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, 423, 752–756.
- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, 22, 464–471.
- Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, 51, 333–342.
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14, 391–396.
- Sorkin, R. D., Hays, C. J., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review*, 108, 183–203.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46, 225–245.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, 35, 3053–3069.
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79, 471–491.
- Yin, R. K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81, 141–145.