Uncovering Gender Discrimination Cues in a Realistic Setting

Nicolas Dupuis-Roy
Département de Psychologie, Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal, QC, Canada

Isabelle Fortin
Département de Psychologie, Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal, QC, Canada

Daniel Fiset
Département de Psychologie, Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal, QC, Canada

Frédéric Gosselin
Département de Psychologie, Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal, QC, Canada

Which face cues are we using for gender discrimination? Few studies have tried to answer this question and all of them suffer from poor external validity, using only a small set of grayscale stimuli, often distorted, and presented a large number of times. Here, we reassessed the importance of facial cues for gender discrimination in a more realistic setting. We applied *Bubbles*—a technique that minimizes bias toward specific facial features and does not necessitate the distortion of stimuli—to a set of 300 color photographs of Caucasian faces, each presented only once to 30 participants. Results show that the region of the eyes and the eyebrows—probably in the light-dark channel—is the most important facial cue for accurate gender discrimination; and that the mouth region is driving fast correct responses (but not fast incorrect responses)—the gender discrimination information in the mouth region is concentrated in the red-green color channel. Together, these results suggests that, when color is informative in the mouth region, humans use it and respond rapidly; and, when it's not informative, they have to rely on the more robust but more sluggish luminance information in the eye-eyebrow region.

Keywords: gender discrimination, color vision, bubbles, classification image, face perception

Introduction

Which face cues are we using for gender discrimination? Up until now, the small body of studies on this topic has highlighted the importance of the eyes, the eyebrows, the jaw and the face outline (e.g., Brown & Perrett, 1993; Yamaguchi et al., 1995; Russell, 2003, 2005; Nestor & Tarr, 2008, submitted). Using *Bubbles*, Schyns, Bonnar and Gosselin (2002; see also Gosselin & Schyns, 2001) found that relatively coarse eye and mouth information (5.62-22.5 cycles per face width for a face width subtending about 4 cycles per degree of visual angle) were significantly correlated with gender discrimination in humans. Relatedly, the distance between the brows and the upper eyelid was identified as the most reliable relational cue to gender in facial images (Campbell, Benson, Wallace, Doesbergh & Coleman, 1999; Burton, Bruce & Dench, 1993). Experiments investigating the role of pigmentation cues showed that human observers could rely on chromatic information—mostly on the red-green axis—to categorize gender especially when minimal discriminative shape information were revealed (Bruce & Langthon, 1994; Hill, Bruce & Akamatsu, 1995; Tarr, Kersten, Cheng & Rossion, 2001; Tarr, Rossion & Doerschner, 2002). The regions surrounding the eyes and the mouth were also found to be the most determinant chromatically (Nestor & Tarr, submitted).

All the studies cited above suffer from at least one of the following three major limitations that decrease their external validity. First, all of them—except Gosselin and Schyns (2001), Schyns et al. (2002), and Nestor & Tarr, 2008—manipulated specific features and regions of the face with techniques such as morphing and caricaturing. These manipulations could have distorted the natural characteristics of authentic faces. Moreover, selective manipulation of these features might have biased the results toward a limited sample of all the facial information available. Second, the face stimuli used in all of these studies—except the studies performed by Tarr and colleagues—were grayscale pictures or they were controlled for different aspects (e.g., hair and ears removed, no makeup). In fact, the skin and hair reflectance properties of males and females differ (makeup only exaggerates these spectral dimorphism—Russell, 2003) and, as we have mentioned above, human observers can use these differences reliably. Third, all of these studies—except Nestor & Tarr, 2008—used a small set of faces that needed to be shown many times to each participant. This context is

doi:10.1167/5.1.1 Received January 1, 2005; published February 18, 2005 ISSN 1534-7362 © 2005 ARVO

likely to have promoted perceptual learning of the faces. Therefore, the results might reflect the peculiarities of the stimulus set rather than general characteristics of gender dimorphism. In fact, the repetition of the same face identity allows the subject to use a face identification strategy rather than a gender discrimination strategy. This may have artificially increased the role of eye region, a potent feature for face recognition (Sekuler, Gaspar, Gold, & Bennett, 2004; Gosselin & Schyns, 2001; Schyns et al, 2002).

Here, we reassess the importance of facial cues for gender discrimination in a more realistic setting: We apply *Bubbles*—a technique that minimizes bias toward specific facial features and does not distort stimuli—to a set of 300 color images of Caucasian faces that were presented only once to 30 participants.

Methods

Subjects

Thirty students from the University of Montreal and McGill University were recruited to participate to the experiment. All participants had normal, or corrected to normal vision. Informed consent was obtained before the beginning of the experiment and a monetary compensation was provided.

Stimuli

Stimuli were generated from 300 color images of Caucasian faces (150 females), chosen on Internet with the intent of ecological representativity. The only other characteristics required for selection were a clear gender membership, a neutral expression and a frontal view. Thus, no special attention was paid to lighting, file format, image size, age of depicted individual, etc.. Subsequent transformations applied on the images were also kept to a minimal. A series of rotations, scalings and translations were applied to the faces to minimize the square of the distance between handpicked landmarks around the eyes (four landmarks each), the eyebrows (two landmarks each), the nose (four landmarks) and the mouth (four landmarks); the average interpupil distance was 40 pixels (1.03 deg of visual angle). Note that these transformations are linear and therefore do not modify any relations between features. Six instances of the resulting face images are shown on Figure 1a.

Stimuli were created by sampling the face images by presenting them behind an opaque mask punctured by an adjustable number of randomly located one-pixel holes (henceforth called '*bubble mask*') and smoothed with a Gaussian kernel having a full-width half maximum of 9.42 pixels (0.24 deg of visual angle). The result, illustrated in Figure 1c, is a partly revealed face on a mid-gray background.

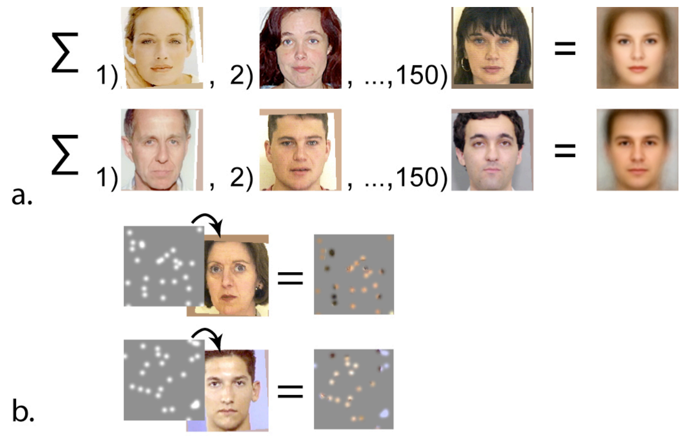

Figure 1. a) Three women and three men from our face database; and the average of all 150 women and 150 men from our face database. b) A stimulus are generated by a overlaying an opaque gray mask punctured by a number of randomly located Gaussian apertures on a face.

Apparatus

The experimental programs were run on a Macintosh G4 in the Matlab environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented on a Sony Trinitron monitor (1024 x 768 pixels at a refresh rate of 85 Hz). We determined the relationship between RGB values and luminance levels (measured with a Samsung SyncMaster 753 df photometer) for each color channel independently; the three bestfitted "gamma" functions were used in the computation of image statistics. Participants were seated in a dim ambient-lighted room at a distance of approximately 75 cm from the computer monitor.

Procedure

Each participant was submitted to 300 trials and, importantly, each trial involved a different face. The presentation order of the 300 faces was randomized. In a given trial, one stimulus—a sparsely sampled face—appeared at the center of computer monitor and remained there until the participant had indicated the gender of the stimulus by pressing a labelled keyboard key. No feedback was provided. The number of bubbles per image was adjusted on a trial by trial basis to maintain performance at 75% correct using QUEST (Watson & Pelli, 1983).

Results and discussion

Participants used an average of 27.06 bubbles and responded correctly on 74.74% of the trials. The average response time was 1.63 sec. The correlation between response time and accuracy was -0.1216 (p<0.001). There was a slight bias toward responding "man" (52.18% of the trials, p<0.01) rather than "woman". No difference was observed between female and male participants (51.58% and 52.72%, *ns*).

Linear Classification Image Analyses

To uncover which facial cues led more often to accurate or faster correct gender discrimination, we performed three least-square multiple linear regressions: one between discrimination accuracies (predictive variable) and bubble masks (explanatory variable); another between quartiles of response time on correct trials and bubble masks[1]; and, a final one, between quartiles of response time on incorrect trials and bubble masks. The outcome of these regressions are three 128 by 128 planes of regression coefficients which are called classification images (Eckstein & Ahumada, 2002; Gosselin & Schyns, 2004). We summed classification images across participants and smoothed the resulting group classification images with a Gaussian kernel having a full-width half maximum of 16.48 pixels. The statistical analysis was restricted to the area of the classification images that could contain face information; the complementary area, which was irrelevant to the task at hand, was used to estimate the mean and the standard deviation of the null distribution and to transform the smooth classification images into Z-scores. Any significant positive local divergence from uniformity in our smooth group classification images would indicate that the corresponding part of the stimuli led to more accurate responses, faster correct responses, or faster incorrect responses. We therefore conducted one-tailed Pixel tests (Chauvin, Worsley, Schyns, Arguin & Gosselin, 2005) on the Z-scores smooth group classification images transformed into Z-scored (S_r = 3469; for accuracy: Z_{crit} = 3.7 and Z_{max} = 6.48; for response time: Z_{crit} = 3.5 and Z_{max} = 4.04; $p < .05$). The statistical threshold provided by this test corrects for multiple comparisons while taking the spatial correlation inherent to our technique into account.

Figure 2 displays the average women (column 1) and men (column 2) overlaid with a contour-plot representation of the accuracy and correct response time classification images. Nothing reached statistical significance in the incorrect response time classification image. The colored pixels enclosed by the dotted black lines are statistically significant: the region of the eyes and eyebrows lead to more accurate and faster correct gender discrimination; this eye-eyebrow region is wider and more bilateraly distributed in the correct response time classification image (row 1) than in the accuracy classification image (row 2); and facial cues leading to fast correct responses also included the mouth region as well as the space between the mouth and the nose. That the mouth is significantly correlated with correct response time but is correlated neither with incorrect response time, nor with accuracy might seem puzzling. Close examination of the data shows that the relatively few correct answers given when the mouth was revealed led to very short reaction time, which suggests that the mouth would become significantly correlated with accuracy if more data were collected.

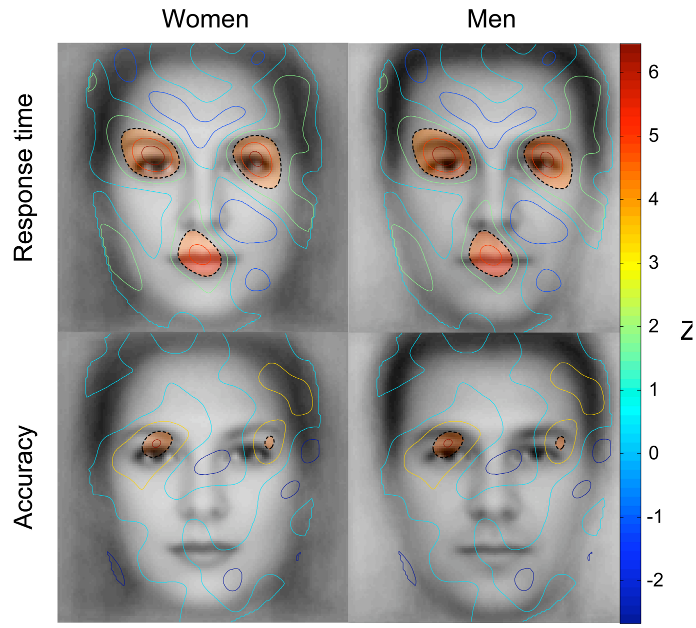

Figure 2. Displays the average men (column 2) and women (column 1) superimposed on a contour plot of classification images derived from accuracy (row 1) and response time (row 2). The colored pixels enclosed by the dotted black lines are statistically significant (p<.05).

Beyond Linear Classifications Images

The linear classification image analyses confirmed that the eye-eyebrow region contains the most important cues for gender discrimination. However, they do not allow to identify more precisely the nature of these reliable cues, at least more directly. For example, we could wonder if these cues are mostly red-green pigmentation cues, as proposed by Tarr and colleagues? It's not so much a limit of the methodology than a limit of the search space we chose to explore—image location. In fact, Nestor and Tarr (submitted) have used classification images to probe the use of color directly during gender discrimination. On each one of 20,000 trials, color noise was added to the same androgynous morph and participants had to decide whether it looked more like a man or a woman. If we cannot address the color question directly, we can provide—based on the 300 faces of our face set—image statistics about the discriminative color information that was available within the eye-eyebrow region.

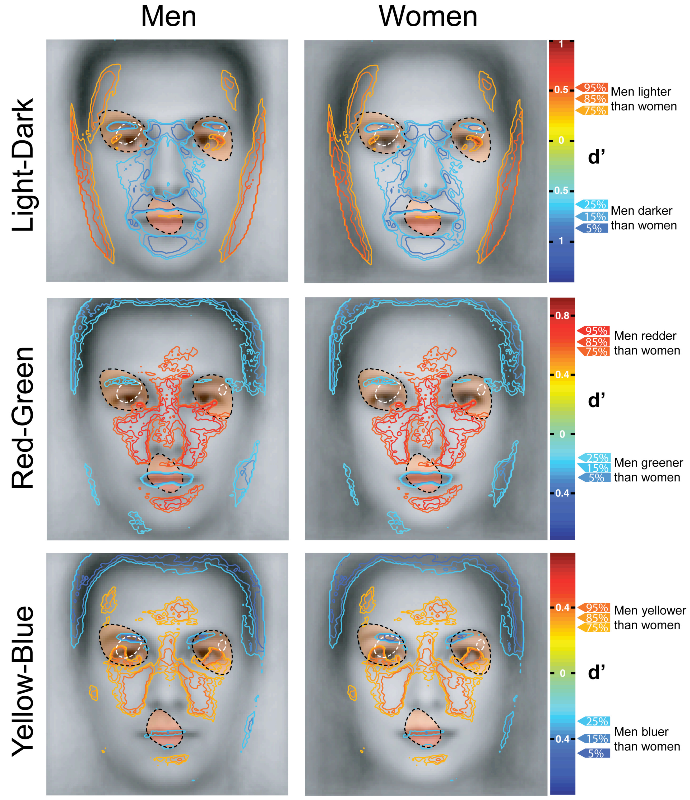

Figure 3. Contour plots of the color maps superimposed to the average man (column 1) and woman (column 2). Dotted lines define clusters significantly correlated with accurate (white) and correct fast responses (black). The contour plot summarizes the spatial modulation of available information (*d*'s) in the dark-light (row 1), red-green channel (row 2) and yellow-blue (row 3) channels. The color-labelled lines of isovalued *d*'s correspond to percentile 95%, 85%, 75%, 25%, 15%, and 5%.

We converted these face images to Lab color space because its channels represent perceptually relevant color opponent processes: L corresponds to the light-dark process, a to the red-green process, and b to the yellow-blue process. Then, we computed d' on each pixel of the three Lab channels—we will call the resulting d' planes color maps. This metric could be interpreted as the information available in a given pixel of a given color map to discriminate the gender. More specifically, a pixel's d' is the distance (in standard deviation units) between the mean of the distribution of this pixel's value for male faces and the mean of the distribution of this pixel's value for female faces. The three color maps are represented as contour plots in Figure 3. Color lines delimit isovalued d' corresponding to percentiles of 95%, 85%, and 75%. Warm colors were used for regions where men are lighter, redder or yellower than women; and cold colors for regions where men are darker, greener or bluer than women. To help with interpretation, the contour plots were placed over an image of the average men (column 1) and women (column 2). Thick dotted lines were added to delineate the significant regression coefficients found in the accuracy (white) and the correct response time classification images (black).

The light-dark color map depicts the information that has been mainly investigated in the litterature so far. It shows the availability of prominent gender cues in the temporal side of the brows and the eyes, over the upper lip, and under the commissure of the chin and the lower lip (Russell, 2003, 2005). Note also the luminance information located on the face outline near the cheeks. On average, this channel has higher d's than the other color channels (mean d's: light-dark = 0.36, red-green = 0.27, yellow-blue = 0.21). This set of informative features overlap substancially with the features found in the accuracy and correct response time classification images. The most informative pixels in the red-green color map—the second most informative color channel—are localized on the lips but are also distributed on the maxilla region and near the chin-lower lip commissure. The upper lip is a feature also found in the correct response time classification image. In comparison, the yellow-blue channel contains less information allowing to distinguish males from females. The most informative yellow-blue cues are clustered on the temporal sclera, on the nasal side of the brows and on the outer portion of the hair. None of these features is found in the classification images.

	Eyes	Brows	Eyelids	Eyes, brows and eyelids
Observed Accuracy	0.7216	0.7329	0.6897	0.7626
Predicted accuracy	0.6082	0.7471	0.6612	0.8216
N	194	87	307	269

Table 1. The first row shows the mean accuracy observed when areas are revealed separately (columns 1-3) and together (column 4). The second row indicates the average accuracy predicted from linear regression. The last row displays the number of trials that were used to compute these statistics.

Another cue that has already been targeted as the one of the most discriminative information for gender categorisation is the eyelid-brow distance (Burton, Bruce & Dench, 1993; Campbell et al., 1999), i.e. the distance between the center of the upper eyelid and the center of the bottom part of the eyebrow. If the participants used this cue they needed, in order to us it, to see part of the eye, the eyelid and the brow together. Therefore, the performance observed in the trials in which these regions were presented together (see Table 1, first row) should be higher than the performance predicted by the linear combination of these regions presented individually with the appropriate weights from the accuracy classification image (see Table 1, second row).

In fact, predictions made from the accuracy linear regression explains the observed the performance observed when the eye, the eyelid and the brow are seen together. Moreover, image statistics computed on the 300 faces from our database indicate that this relational cue provide little discriminative information: the d' of the eye-eyelids distance—measured from handpicked landmarks—is 0.91. In sum, these results do not support the use of the eyelid-brow distance in our experiment. Further analyses would be required to assess the use of other distance cues. However, Nestor and Tarr (2008) performed a similar analysis on all pairwise conjunctions between the forehead, the eyes, the ears, the upper and lower part of the nose, the cheeks, the mouth, and the chin, and failed to found evidence for nonlinear use of information during their gender discrimination task.

Conclusion

Which face cues are we using for gender discrimination? All studies that have previously attempted to answer this question suffered from poor external validity. They typically manipulated specific features and regions of the face with techniques such as morphing and caricaturing. These manipulations probably altered the natural characteristics of faces, and biased the results. We sampled unaltered face photographs with minimum bias by presenting them behind mid-gray opaque masks punctured by a number of randomly located Gaussian apertures. Furthermore, the face stimuli that have been used typically in gender discrimination experiments are grayscale photographs, normalized and controlled for different aspects. Our results can be considered as more representative of genuine gender discrimination because our face stimuli were real-life colour photographs and, therefore, were not (artificially) controlled for luminance, chrominance, background, hair and makeup. Previous studies on facial gender discrimination cues used a small set of faces that needed to be shown many times to each participant; therefore, the results might reflect the peculiarities of small stimulus sets overlearned by participants rather than general characteristics of gender dimorphism. We used a set of 300 face photographs that were presented only once to each one of our 30 participants.

To uncover which facial cues led more often to accurate or faster correct gender discrimination, we performed three classification image analyses: on accuracies, on correct response times, and on incorrect response times. The accuracy classification image confirmed that the eye-eyebrow region is the most important cue for gender discrimination (e.g., Brown & Perrett, 1993; Yamaguchi et al., 1995; Schyns, Bonnar & Gosselin, 2002). Linear predictions made on a sub-set of our trials showed that participants did not use the eyelid-brow distance information, a distance cue that Burton, Bruce and Dench (1993) and to Campbell et al. (1999) stated was one of the most reliable for gender discrimination. In fact, we discovered that the eyelid-brow distance has a small signal-to-noise ratio for gender discrimination.

Image statistics computed on the *Lab* channels showed that the highly discriminative information contained in the eye-eyebrow area is mostly concentrated in the light-dark channel. This suggests that humans discriminate face gender based on a linear combination of luminance cues within the eye-eyebrow region. There is no inconsistency between our results and Tarr and colleagues' results about the important role of color in face gender discrimination (Tarr, Kersten, Cheng & Rossion, 2001; Tarr, Rossion & Doerschner, 2002; Nestor & Tarr, 2008, submitted). They showed that participants relied on pigmentation cues (especially from the red-green channel) when minimal or no luminance information is available. Similarly, Yip and Sinha (2002) showed that color cues play a role in face identification when shape attributes are degraded. Yip and Sinha proposed that the contribution of color may lie not so much in providing diagnostic cues to identity as in aiding low-level image-analysis processes such as segmentation; and the same could be proposed about face gender discrimination.

This being said, the incorrect and correct response time classification images suggest a more ubiquitous role for color in face gender discrimination. The mouth region is significantly correlated with correct fast responses (but not with incorrect fast responses) and the most discriminative information in the mouth region is concentrated in the red-green channel. This suggests that humans do use chromatic cues for discriminating face gender: When it's informative, they use it and respond rapidly (for evidence that color is perceived faster than shape, see Moutoussis and Zeki 1997a; 1997b; Holcombe & Cavanagh, 2001); when it's not, they have to rely on the more robust and more sluggish luminance cues. The infero-temporal cortex, which is involved in both face perception and color perception (Clark et al., 1997; Edwards, Xiao, Keysers, Foldiak & Perrett, 2003), provides the ideal *locus* for such a dual strategy.

Acknowledgments

This research was supported by FQRNT scholarship awarded to Nicolas Dupuis-Roy and Isabelle Fortin, and by NSERC and NATEQ grants awarded to Frédéric Gosselin.

Commercial relationships: none.
Corresponding author: Nicolas Dupuis-Roy
Email: nicolas@dupuis.ca
Address: Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, QC, H3C 3J7, Canada.

Footnotes

[1]For the regression on accuracy, we subtracted the mean of the bubble masks that led to an incorrect response from the mean of the bubbles masks that led to a correct response. And, for the regression on response time, we summed 1.5 times the mean of the bubble masks that led to a correct response and to a response time in the fastest quartile, 0.5 times the mean of the bubble masks that led to a correct response and to a response time in the second quartile, -0.5 times the mean of the bubble masks that led to a correct response and to a response time in the third quartile, and -1.5 times the mean of the bubble masks that led to a correct response and to a response time in the slowest quartile. Prior to these computations, every bubble mask was divided by the number of one-pixel holes it contained to give equal weight to all bubble masks.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Brown, E., & Perrett, D. I. (1993). What gives a face its gender. *Perception, 22*(7), 829-840.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., et al. (1993). Sex-discrimination: How well do we tell the difference between male and female faces. *Perception, 22*(2), 131-152.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. Perception, 23, 803–822.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22*, 153–176.

Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception, 28*(4), 489-504.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision, 5*(9), 659-667.

Clark, V. P., Parasuraman, R., Keil, K., Kulansky, R., Fannon, S., Maisog, J. M., et al. (1997). Selective attention to face identity and color studied with fMRI. *Human Brain Mapping, 5*(4), 293-297.

Edwards, R., Xiao, D. K., Keysers, C., Foldiak, P., & Perrett, D. (2003). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *Journal of Neurophysiology, 90*(2), 1245-1256.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261-2271.

Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the sex and race of faces: The role of shape and colour. *Proceedings of the Royal Society B: Biological Sciences, 261*, 367–373.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience, 4*(2), 127-128.

Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London Series B-Biological Sciences, 264*(1380), 393-399.

Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London Series B-Biological Sciences, 264*(1387), 1407-1414.

Nestor, A., & Tarr, M. J. (2008). The segmental structure of faces and its use in gender recognition. *Journal of Vision, 8*(7):7, 1–12, http://journalofvision.org/8/7/7/, doi:10.1167/8.7.7.

Nestor, A. & Tarr, M. (submitted). Recognition of human faces using color. *Psychological Science.*

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. Perception, 32, 1093–1107.

Russell, R. (2005). Face pigmentation and sex classification [Abstract]. Journal of Vision, 5(8):983, 983a, http://journalofvision.org/5/8/983/, doi:10.1167/5.8.983.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402-409.

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.

Tarr, M.J., Kersten, D., Cheng, Y., & Rossion, B. (2001). It's Pat! Sexing faces using only red and green [Abstract]. *Journal of Vision, 1*(3):337, 337a, http://journalofvision.org/1/3/337/, doi:10.1167/1.3.337.

Tarr, M. J., Rossion, B., & Doerschner, K. (2002). Men are from Mars, women are from Venus: Behavioral and neural correlates of face sexing using color [Abstract]. *Journal of Vision, 2*(7):598, 598a, http://journalofvision.org/2/7/598/, doi:10.1167/2.7.598.

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptative psychometric method. *Perception & Psychophysics, 33*(2), 113-120.

Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception, 24*(5), 563-575.

Yip, A. W. & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 31*, 995-1003.