

Is it a he or a she? Behavioral and computational approaches to sex categorization

Gwenaël Kaminski · David Méary ·
Martial Mermillod · Edouard Gentaz

Published online: 4 May 2011
© Psychonomic Society, Inc. 2011

Abstract Can people categorize the sex of neonate faces? Our experiment tested the sex categorization of neonate faces by adult participants. We used a set of 120 Caucasian faces (adults and 4-day-old neonates) that were presented just once to a large sample of participants. A computational model of low-level visual processing, based on Gabor filters, was used to explore the relation between spatial-frequency information and sex categorization. The results showed that participants were able to categorize the sex of the faces, but were less accurate with neonate ($d' = 0.36$, $\beta = -.97$) than with adult ($d' = 3.02$, $\beta = -.93$) faces. Moreover, faces were more frequently categorized as boys' than girls' faces. The computational model suggests that specific spatial-frequency channels carry most of the useful information for the categorization task. Overall, the findings reveal that subtle differences in neonate facial structure were enough to allow the sex categorization of neonate faces, although accuracy was low in both adults and the computational model of low-level visual processing.

Electronic supplementary material The online version of this article (doi:10.3758/s13414-011-0139-1) contains supplementary material, which is available to authorized users.

G. Kaminski (✉) · D. Méary · E. Gentaz
LPNC, CNRS, UMR 5105, Université Pierre Mendès France,
38040 Grenoble, France
e-mail: Gwenael.Kaminski@upmf-grenoble.fr

M. Mermillod
Clermont Université, Université Blaise Pascal,
and CNRS, UMR 6024,
Clermont-Ferrand, France

M. Mermillod
Institut Universitaire de France (IUF),
Toulon, France

Is it a male or a female? Answers to this question are rapid and accurate when viewing adult or child faces (Wild et al., 2000) but are only slightly better than chance when judging infant faces (Porter, Cernoch, & Balogh, 1984; Round & Deheragoda, 2002). Why is sex categorization of infant faces so difficult?

In adult faces, males and females have been found to differ in terms of physical structure, such as the absolute and relative distances between facial features (Burton, Bruce, & Dench, 1993). Among the facial features that contribute to sex categorization, other than the nose (Brown & Perrett, 1993), the eye and eyebrow regions are thought to be most important for categorization (Dupuis-Roy, Fortin, Fiset, & Gosselin, 2009). In addition, people may use adult facial contrasts (between the eyes and/or mouth and the rest of the face) to determine the sex of a face (Russell, 2009), and it has been suggested that low-spatial-frequency information may be more useful for sex categorization than high-spatial-frequency information (see, e.g., Abdi, Valentin, Edelman, & O'Toole, 1995; Sergent, 1986). Previous studies have also demonstrated that additional cues, such as hairstyle and clothing, increase sex categorization performance (e.g., MacRae & Martin, 2007). Based on some or all of these indices, categorizing adult faces according to sex is straightforward.

In children's faces, however, few studies have focused on sex categorization. The plausible reason is that there are very few differences in facial skeletal structure between prepubescent boys and girls but numerous differences in adult faces (see, e.g., Enlow, 1982). Nevertheless, it seems clear from a perceptual standpoint that there is sufficient information in children's (Cheng, O'Toole, & Abdi, 2001; Wild et al., 2000) and neonates' (Porter et al., 1984; Round & Deheragoda, 2002) faces to allow for accurate sex

categorization. The apparent contradiction between skeletal and behavioral information suggests that adults may use other perceptual cues instead of facial traits to assess the sex of a face (e.g., holistic information or different spatial-frequency [SF] channels), or that adults' ability to categorize sex in infants' faces has been overestimated. There are reasons to believe that the two previous studies on neonate sex categorization (Porter et al., 1984; Round & Deheragoda, 2002), which showed that adult performance is only slightly better than chance, suffered from various experimental limitations that may have affected or biased the results. These include small neonate facial samples (28 and 30 full-term infants in Porter et al., 1984, and Round & Deheragoda, 2002, respectively), faces with sex-stereotyped cues (in both studies), and faces belonging to various ethnic groups (also in both studies). Clearly, a better understanding of adult categorization of infant faces will require carefully controlled measures of sex categorization performance.

To advance our understanding of the adult ability to categorize the sex of neonate faces, we used a behavioral and computational approach. In doing so, we tested whether adults are able, in a carefully controlled setting (consisting of a large set of facial stimuli, with all sex-stereotyped cues removed, and a single ethnic group for participants and facial stimuli), to categorize the sex of unfamiliar neonate faces. Because SF information was not available in the skeletal analysis, and because it has been suggested that low SFs are more useful than high SFs in performing sex categorization in adult faces (Abdi et al., 1995), we also examined the role of SF information (Gaspar, Sekuler, & Bennett, 2008). To this end, we used a computer vision model derived from the biological properties of the human perceptual system (De Valois & De Valois, 1988) that combined Gabor wavelet filtering (Mermillod, Bonin, Mondillon, Alleysson, & Vermeulen, 2010; Mermillod, Vuilleumier, Peyrin, Alleysson, & Marendaz, 2009a) and an exemplar-based categorization model (Nosofsky, 1988).

Method

Participants

A group of 76 adults participated in the experiment (32 men, 44 women; mean age = 21.7 ± 0.9 years, min–max = [20–25]). All of the participants were students at Grenoble University. All were Caucasian, and none had children. The participants were selected prior to testing via a questionnaire to ensure that none had any extensive past exposure to infants' faces (Kuefner, Macchi Cassia, Picozzi, & Bricolo, 2008).

Stimuli and procedure

The stimuli consisted of 120 photographs of Caucasian faces—100 full-term neonates (50 boys, 50 girls; mean age = 101 h) and 20 adults (10 men, 10 women; mean age = 22.3 years, min–max = [18–25])—without outer facial features (see Fig. 1). We decided to use adult in addition to neonate faces to ensure that our experimental paradigm was working correctly. Because we expected sex categorization of the adult faces to be nearly perfect (Wild et al., 2000), the adult categorization task served as our baseline condition. None of the faces were familiar to the participants. Photos of the neonates' faces were taken in a nursery. The image areas and luminosity were equalized, and the faces were pasted on a uniform gray background (580×580 pixels). Participants were asked to categorize the sex of all 120 grayscale faces. Because the two previous studies showed that neonate sex categorization is difficult, we chose a long stimulus exposure time (relative to traditional visual experiments) to improve possible categorization. The faces were presented for 5 s, one at a time, on a computer screen. The participants then had another 5 s to reach a decision (boy or girl). More detailed information about the facial stimuli and procedure are included in the electronic supplementary information (section S1)

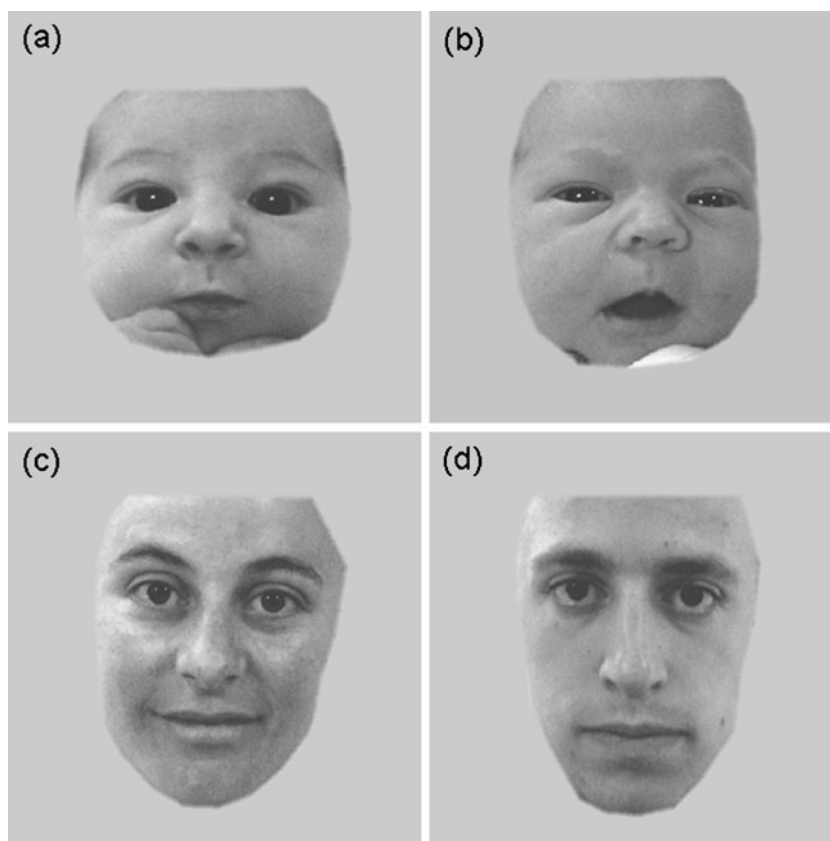
Gabor wavelet filtering

To examine SF information and its effect on sex categorization, we used a bank of 56 Gabor wavelet filters (GWFs). The model was made up of seven SFs (one octave per SF channel: 59.4, 29.7, 14.8, 7.4, 3.7, 1.8, and 0.9 cycles per image) and eight orientations (0 , $\pi/8$, $2\pi/8$, $3\pi/8$, $4\pi/8$, $5\pi/8$, $6\pi/8$, and $7\pi/8$ radians, with 0 at horizontal in spectral domain), in line with neurofunctional descriptions of the primary visual cortex (De Valois & De Valois, 1988). Each face was therefore encoded as a 56-element vector, corresponding to the magnitude of the 56 responses provided by the GWF (details in Mermillod, Vermeulen, Lundqvist, & Niedenthal, 2009b). We performed a cluster analysis of the 120×56 element matrix representing our stimulus set to compute the Euclidean distance between two faces. Then, for each neonate's sex category (girl or boy), based on the exemplar model (the perceptual distance between one individual and all other individual exemplars), we measured the median Euclidean distance between each face and the model (for more details, see section S2 of the online supplement).

Statistical analysis

Our task is analogous to a signal detection problem. Signal detection estimates of sensitivity d' , as well as a likelihood

Fig. 1 Examples of stimuli. We used both neonate (**a**, **b**) and adult (**c**, **d**) faces, as well as female (**a**, **c**) and male (**b**, **d**) faces



criterion, were used to determine whether participants were able to distinguish between male and female faces. We compared the d' measure with the reference score (for no sex detection, $d' = 0$), and significance was assessed with a t test, which we also used when comparing criteria with the no-bias value ($\beta = 1$). We analyzed signal detection estimates for each stimulus age (neonate or adult) and then assessed the effects of stimulus sex on judges' choices in neonate selection, by logistic regression, with judge and stimulus as random effects. Finally, we investigated the effects of stimulus sex, stimulus age (neonate or adult), and their two-way interaction on the entire set of 56 Gabor filters and on each of the seven SFs, using MANOVA. Nonparametric tests were used to compare average Euclidean distances. Statistical analyses were conducted using SAS.

Results

Participant responses

Participants were able to categorize the sex of both adult ($d' = 3.02 \pm 0.32$; $p < .0001$) and neonate ($d' = 0.36 \pm 0.25$; $p < .001$) faces. However, adults' faces were more accurately categorized than those of neonates (paired t test = 64.67, $p <$

.0001, $\eta^2 = .98$). In addition, boys' faces were categorized significantly more correctly than girls' faces (odds ratio = 1.29; $F(1, 7309) = 23.3$, $p < .0001$; see Table 1 and supplemental section S3). We noted a small response bias in neonate faces, in that judges responded "boy" more frequently than "girl" ($\beta = .97 \pm .11$; $p = .024$). That is, they erred on the side of misclassifying girl faces as boy faces.

Gabor wavelet filtering

Incorporating all faces and SFs, the MANOVA analysis revealed significant main effects of both stimulus age [$F(56, 61) = 20.59$, $p < .0001$, $\eta^2 = .95$] and stimulus sex [$F(56, 61) = 2.31$, $p < .001$, $\eta^2 = .68$]. The two-way interaction was also significant [$F(56, 61) = 2.29$, $p < .001$, $\eta^2 = .68$]. Contrast comparisons showed a stimulus sex effect in the adult population [$F(56, 61) = 2.52$, $p < .001$, $\eta^2 = .70$], but not in the neonate population [$F(56, 61) = 1.21$, $p = .24$, $\eta^2 = .52$]. Thus, the model did not perform efficient neonate sex categorization when using the totality of SF information. However, when our analysis was carried out on each SF independently, neonate sex was categorized significantly using the highest SF (59.4 cpi) and the two lowest SFs (1.8 and 0.9 cpi), suggesting that diagnostic cues for neonate sex categorization may occur at specific SFs.

Table 1 Sex categorization using both behavioral (“Judge”; mean \pm SD) and computational (Euclidean distance “Eucl. Dist.”; median with 1st and 3rd quartiles) approaches. We placed each neonate item in one

of three categories, depending on its categorization score: correct, indistinguishable, or false categorization (cat.). We used a total of 100 neonate faces

Categorization	Total	Boys’ Faces			Girls’ Faces		
		<i>N</i>	Judge	Eucl. Dist.	<i>N</i>	Judge	Eucl. Dist.
Correct cat.	54	32	.74 \pm .11	2.25 [2.03–2.89]	22	.72 \pm .10	1.76 [1.64–2.12]
Indistinguishable cat.	16	6	.49 \pm .04	2.35 [2.20–2.45]	10	.50 \pm .04	1.79 [1.62–1.91]
False cat.	30	12	.29 \pm .09	2.20 [2.02–3.49]	18	.32 \pm .10	2.24 [1.64–2.65]
All cat.	100	50	.60 \pm .22	2.27 [2.04–3.08]	50	.53 \pm .20	1.82 [1.64–2.23]

Details on the GWF analyses are shown in supplemental section S4.

Euclidean distance

In neonate faces, we did not find any sex effect on the Euclidean distance between each face and the exemplar model ($z = -0.90$, $p = .37$). However, we found that the median Euclidean distance was smaller in the girl category than in the boy category ($z = -4.78$, $p < .001$; see Table 1 and Fig. 2). In other words, girls’ and boys’ faces have a specific variance in each intrasex category. We also found asymmetry in sex categorization: Any given girl had a smaller median Euclidean distance to all of the other girls’ faces than to any of the boys’ faces ($z = 6.11$, $p < .001$), while any given boy also had a smaller median Euclidean distance to all of the girls’ faces than to any of the other

boys’ faces ($z = 6.04$, $p < .001$). In order to provide a better understanding of these statistical distributions, we performed a principal components analysis on the GWF output (Fig. 2). The figure shows the higher perceptual variability for male than for female neonates, but also that the overlap between the two sex distributions is high.

Discussion

Is the baby a boy or a girl? Our research shows that adults are able to categorize the sex of 4-day-old human faces, confirming the findings shown in two previous studies (Porter et al., 1984; Round & Deheragoda, 2002). Moreover, because we used a large sample of participants and of standardized facial images, the research showed additional results worth emphasizing. Boys’ faces were more correctly categorized than girls’ faces. In addition, as expected, adult sex categorization was more accurate than neonate sex categorization. The latter result may be interpreted in at least two ways. First, it is conceivable that neonate sexual facial cues are more subtle and easily missed than those of adults, making sex categorization less accurate. This is consistent with previous findings (e.g., Enlow, 1982) showing that sexual dimorphism increases with age. A second explanation may be, as Cheng et al. (2001) suggested, that people judge the sexes of adult and infant faces using feature sets derived from the appropriate age category, rather than by applying features derived from another age category or from a combination of age categories. Since the participant age was closer to those of the adults pictured in the facial stimuli than to those of neonates, adult participants were more adept at categorizing adult faces than infant faces, leading to lower performance with neonate face stimuli. Indeed, adults have been shown to perform better when recognizing adult faces than with faces of different ages, resulting in an *other-age effect* (Kuefner et al., 2008; Lamont, Stewart-Williams, & Podd, 2005). Finally, by including the output from the GWF model in our analysis, we found that certain spatial

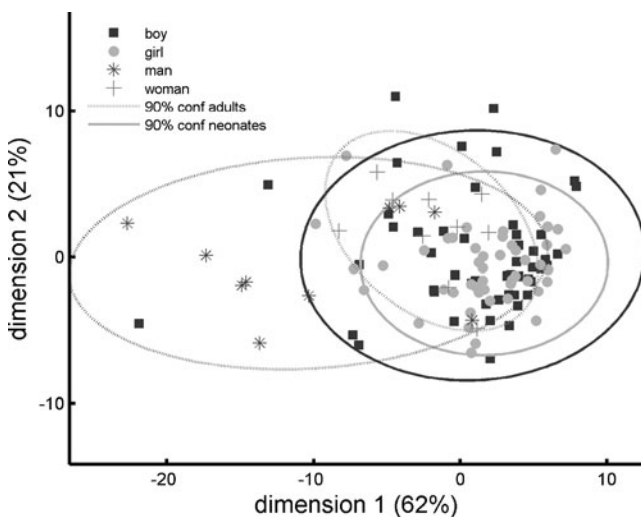


Fig. 2 Illustration, using principal components analysis, of gender and age categorization performed by Gabor wavelet filters. Each face is projected on the first eigenvector (62% of variance, Dimension 1) and the second eigenvector (21% of variance, Dimension 2). The 90% confidence ellipses use the quartiles of the chi-square distributions; solid lines indicate the regions for neonate faces, and dashed lines indicate adult faces

frequencies appear more useful than others in categorizing the sex of neonate faces. This suggests that facial sexual dimorphism may begin with specific visual cues in infants (related to the lowest and highest SF channels), and later become generalized to all SFs in older faces. Further behavioral experiments will need to be conducted to test this hypothesis on human participants.

Two biases that may account for the better sex categorization of boys' faces emerged from our results. First, participants responded "boy" more frequently (the same "male bias" is common in other sex categorization studies—e.g., Wild et al., 2000), favoring correct categorization in the boys' group. Second, the principal components analysis revealed that the distribution of girls' faces along the two main dimensions occurred within that of the boys' faces, which had greater variance. This asymmetry in variability and the overlap between the two categories is the type of statistical distribution that leads to categorization asymmetries in artificial and biological neural systems (French, Mareschal, Mermillod, & Quinn, 2004; Mermillod et al., 2009b). Neonate faces close to the central values have similar SF content, and discrimination of sex based on SF may be more difficult than for those faces located farther from the neonate group (Baudouin & Gallay, 2006). Boys' faces have a greater variance and stuck out of the neonates' group more frequently, which may have improved sex categorization in the group. Such a hypothesis is in line with Valentine's (1991) "face space" model, which considered the representation of a face to be a point in a multidimensional space. The dimensions of this space are the featural properties (distance between eyes, hair color . . .) used to encode faces, and it has been suggested that sex is another dimension of the multidimensional space (Baudouin & Tiberghien, 2002; but see Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000, suggesting that facial sex processing is performed in parallel with the perceptual analysis of facial features). Our data, however, do not allow us to determine whether neonate faces were perceived as variations of a nonsexed face prototype (one high-density facial area, equivalent to the neonate sex area) or whether two sexed face prototypes (two high-density areas, one for each neonate sex category) were close to each other (Baudouin & Gallay, 2006).

SF information, assessed by a Gabor wavelet model, furthers our understanding of sex categorization. Indeed, the computational findings suggest that sex categorization of adult faces should be easier than that of neonate faces, so the results are consistent with our behavioral approach. Our model also revealed that the output from certain filters (the highest and two lowest SFs) alone could be used to predict neonate sex. However, as illustrated in Fig. 2, we must acknowledge that the GWF model, when including all SFs, was not very effective at categorizing the sex of neonate

faces, because most of the faces were located within the area of overlap between the boy and girl groups. Further behavioral research will be needed to investigate whether these specific SF channels on their own are sufficient to perform neonate sex categorization by humans. If they are, since SFs cannot be measured from the biometric data, the use of SFs should provide a reasonable explanation for the sex categorization contradiction between the skeletal (no sex difference) and perceptive (weak but reliable sex categorization) findings in the literature.

In sum, our study pushes the limits of adult ability in facial sex categorization by showing that adults are able to categorize the sexes of 4-day-old neonates. When adults encounter a neonate face, however, they only have a 60% chance of accurately doing so. To grasp the sex cues used by the adult facial processing system, we need to identify the components that favor categorization in neonate faces. This can be achieved by using a masking paradigm, as shown by Dupuis-Roy et al. (2009). Finally, working with expert judges on neonate faces (pediatricians, nurses) may highlight the role that experience plays in sex categorization.

Author Note This research was funded by the National Center for Scientific Research (CNRS) and by grants from the National Research Agency (ANR Family'Air grant to E.G.; ANR Grant BLAN06-2_145908, ANR Grant ANR-06-CORP-019, and an Institut Universitaire de France grant to M.M.). G.K. proposed the project and performed the statistical analyses. G.K., D.M., and E.G. designed and conducted the experiments. G.K. and M.M. performed the Gabor filter analyses. All authors wrote and discussed the results and commented on the manuscript. The study was performed in accordance with the Declaration of Helsinki: It was conducted with the understanding and the written consent of each participant, and was approved by the local ethics committee. All parents gave informed written consent for the limited use of their neonates' pictures. The authors have declared that no competing interests exist. We thank all of the neonates, their parents, the staff members working at the maternity ward of the Clinique Mutualiste in Grenoble, France, and the adult judges who participated in the experiment. Many thanks to Mathieu Gallay and to Benjamin de Vulpillières for their constructive comments on the manuscript.

References

- Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal-component approach. *Perception*, 24, 539–562.
- Baudouin, J.-Y., & Gallay, M. (2006). Is face distinctiveness gender based? *Journal of Experimental Psychology: Human Perception and Performance*, 32, 789–798.
- Baudouin, J.-Y., & Tiberghien, G. (2002). Gender is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 362–365.
- Brown, E., & Perrett, D. I. (1993). What gives a face its gender. *Perception*, 22, 829–840.

- Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, 22, 153–176.
- Cheng, Y. D., O'Toole, A. J., & Abdi, H. (2001). Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding. *Cognitive Science*, 25, 819–838.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
- Dupuis-Roy, N., Fortin, I., Fiset, D., Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision*, 9(2), 10:1–8.
- Enlow, D. (1982). *Handbook of facial growth*. Philadelphia: Saunders.
- French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, 133, 382–397.
- Gaspar, C., Sekuler, A. B., & Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, 48, 2817–2826.
- Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all babies look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 807–820.
- Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition*, 33, 1017–1024.
- Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, 37, 793–805.
- Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *The European Journal of Neuroscience*, 12, 303–310.
- Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, 73, 2522–2531.
- Mermillod, M., Vuilleumier, P., Peyrin, C., Alleysson, D., & Marendaz, C. (2009a). The importance of low spatial frequency information for recognizing fearful facial expressions. *Connection Science*, 21, 75–83.
- Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009b). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, 110, 346–357.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 54–65.
- Porter, R. H., Cernoch, J. M., & Balogh, R. D. (1984). Recognition of neonates by facial-visual characteristics. *Pediatrics*, 74, 501–505.
- Round, J. E. C., & Deheragoda, M. (2002). Sex—Can you get it right? *British Medical Journal*, 325, 1446–1447.
- Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, 38, 1211–1219.
- Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. M. Young (Eds.), *Aspects of face processing* (pp. 17–33). Dordrecht: Kluwer.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43, 161–204.
- Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, 77, 269–291.