

The MR2: A multi-racial, mega-resolution database of facial stimuli

Nina Strohminger^{1,6} · Kurt Gray² · Vladimir Chituc³ · Joseph Heffner⁴ · Chelsea Schein² · Titus Brooks Heagins⁵

© Psychonomic Society, Inc. 2015

Abstract Faces impart exhaustive information about their bearers, and are widely used as stimuli in psychological research. Yet many extant facial stimulus sets have substantially less detail than faces encountered in real life. In this paper, we describe a new database of facial stimuli, the Multi-Racial Mega-Resolution database (MR2). The MR2 includes 74 extremely high resolution images of European, African, and East Asian faces. This database provides a high-quality, diverse, naturalistic, and well-controlled facial image set for use in research. The MR2 is available under a Creative Commons license, and may be accessed online.

Keywords Faces · Psychological instruments

✉ Nina Strohminger
nina.strohminger@yale.edu

¹ Kenan Institute for Ethics, Duke University,
Durham, NC, USA

² Department of Psychology, University of North Carolina,
Chapel Hill, NC, USA

³ Center for Advanced Hindsight, Duke University,
Durham, NC, USA

⁴ Department of Philosophy, Duke University,
Durham, NC, USA

⁵ Titus Brooks Heagins Photography,
Durham, NC, USA

⁶ Yale School of Management, Yale University,
New Haven, CT, USA

The face is the mirror of the mind. A person's face conveys unparalleled information, including psychological state, sex, age, health, and cultural identity, through cues such as muscle activation (Ekman & Friesen, 1976), eye gaze (Baron-Cohen et al., 2001), skin texture (Fink et al., 2001), luminance (Russell, 2003), eye size (Sacco & Hugenberg, 2009), and face shape (Perrett et al., 1994). In real life, we view faces in the highest of definition, and are able to examine these cues—both consciously and unconsciously—to form interpersonal judgments. By contrast, experimental face research typically relies upon low-resolution photographs or computer-generated images. Researchers are generally willing to make this trade-off because they require stimuli that are highly controlled—such as similar gaze and expression—as well as diversity across genders and races. Ideally, however, face databases should be highly controlled, with racial and gender diversity, and portraying a level of detail approximating real life. In this paper, we present such a database, consisting of 74 mega-resolution images of men and woman of European, African, and East Asian ancestry.

Given the wealth of information that faces communicate, it is little surprise that the mind treats the face as special (Yin, 1969; Palermo & Rhodes, 2007; Chalup et al., 2010). Newborns preferentially orient towards face-like stimuli (Goren et al., 1975), and distinct anatomical regions of the brain are devoted to facial processing (Damasio et al., 1990; Kanwisher et al., 1997; Haxby et al., 2000), with faces and facial features evoking a rapid electrocortical response not shared by scrambled faces or other objects (Bentin et al., 1996).

It is generally understood that faces are essential to social cognition (Hugenberg & Wilson, 2013). Situation-specific facial cues can be used to infer an agent's intentions and emotional state (Emery, 2000; Tomasello et al., 2007;

Teufel et al., 2010), and situation-invariant structural features are the basis of a wide range of attributions from personality to attractiveness (Thornhill & Gangestad, 1999; Todorov et al., 2008).

There is high demand for, and use of, facial stimuli in psychological research. According to a survey of the psychological literature using psycINFO, there have been more than 2200 published papers using facial stimuli since the turn of the 21st century. Given the pervasive use of faces in research, and their importance in psychological processing, high-quality databases of facial stimuli are crucial. However, a survey of the available face databases—of which there are many—suggests a need for more modern, naturalistic, and detailed stimulus sets.

Quite a few popular databases are grayscale (Ekman & Friesen, 1976; Bellhumer et al., 1997; Beaupré & Hess, 2005), where they lose much of the information usually recruited for face recognition (Yip & Sinha, 2002). Likewise, the size of images is often small, rarely exceeding dimensions of 600 × 400 pixels and resolutions of 100 pixels per inch (Nordstrøm et al., 2004; Minear & Park, 2004), and in some cases are as small as 250 × 250 pixels (72 ppi) (Righi et al., 2012). Furthermore, most databases are available only in a compressed, lossy format such as JPEG.¹ Image resolution matters for a variety of reasons. With the ever-increasing screen resolutions of computer monitors, it is difficult to use these lower-resolution images without them looking dated or degraded. Researchers may be interested in studying or measuring details of the face, such as eyes or skin texture. Resolution also becomes important when manipulating images, as quality can quickly degrade across successive editing procedures. Most of all, image resolution is important for realism. Humans are highly sensitive to facial information; the more that information is available in the testing stimulus, the more ecological validity it will have.

Many photo sets have a related issue: they comprise faces that are clearly morphed, manipulated, or otherwise artificially generated (Phillips et al., 1998; Perrett et al., 1999; Nosek et al. 2007; Oosterhof & Todorov, 2008). In some cases, this may be desirable or even necessary, as when controlling for precise psychophysical characteristics. However, humans are so exquisitely attuned to facial stimuli that even subtle adjustments are often readily apparent. This can lead to unintended downstream effects, such as the faces taking on a creepy or unsettling appearance (Mori, 1970; Gray & Wegner, 2012). Faces that are largely or

wholly computer generated may be processed differently than those perceived as real (Tinwell et al., 2011; Wheatley et al., 2011). While facial stimuli certainly can be adjusted with advanced software without seeming digitally altered, these precautions have not always been observed in previous databases.

Naturalism in facial stimuli is highly desirable, but stimuli should also be controlled on as many extraneous dimensions as possible. Many extant databases are poorly controlled, with variations in lighting, distance from subject, lens angle, and pose across different faces (Phillips et al., 1998; Minear & Park, 2004; Yin et al., 2006; Milborrow et al., 2010). The result is distracting, clumsy-looking sets with low internal consistency. Clothing, hair styles, makeup, eyeglasses, and jewelry can quickly make an image look dated or fixed in a particular cultural moment (Ekman & Friesen, 1976; Ekman & Matsumoto, 1993; Tottenham et al., 2009). While this problem may not be completely escapable, it can be minimized, and is a necessary step if the researcher wishes subjects to be judging the face, rather than the conditions under which the image was obtained.

Lastly, only a limited number of the available databases provide extensive racial and gender diversity. Sets tend to comprise a single race (Mandal, 1987; Lundqvist et al., 1998; Wang & Markham, 1999; Mandal et al., 2001; Ebner et al., 2010) or just two (Ekman & Matsumoto, 1993; Ma et al., 2015), or a single gender (Eberhardt et al., 2006). Those databases that do contain a mix of races may not have more than one or two stimuli within a single race/sex category (Barrett & Bliss-Moreau, 2009). The use of such sets make generalizing across race and gender, or testing for effects between race and gender, untenable.

The present database may be most directly compared to the Chicago Face Database (CFD; Ma et al., 2015). The CFD is a large, contemporary stimulus set with high resolution. Although there are fewer total faces in the MR2 than the CFD, the MR2 images have a much higher resolution (the images are approximately 200 % larger), has three races rather than two (East Asian as well as European and African), and the images have higher consistency. Consistency is defined as stimuli that are similarly positioned and sized, with little distracting details at the periphery. In many high-speed visual tasks, even small variations in visual angle and details and introduce noise into the data (Healey et al., 1996). In this paper, we provide empirical evidence that subjects can reliably detect the difference in consistency between the two databases.

Given the limitations in extant facial database realism and resolution, consistency and control, and diversity and dating, there is a real and unaddressed methodological need in the field. This paper introduces a database that attempts to address the limitations described above. The images provide a high level of detail and are larger than life-size; they

¹In a lossy format, information is lost whenever the image is opened and modified, or even saved. This means that image quality becomes progressively worse with every adjustment. While lossy formats may be appropriate for endpoint uses, such as presenting web content (where their more efficient file size is an asset), they are not the appropriate choice for preserving the integrity of original images.

Table 1 Means and standard deviations (*in parentheses*) for each measure in the norming study, divided by race and sex category

Sex	Race	<i>N</i>	Attractiveness	Age	Mood	Femininity	Perceived race	Trustworthiness
Female	African	18	3.66 (0.94)	27.51 (5.25)	3.18 (1.38)	4.83 (1.09)	97.00 (0.05)	3.93 (1.25)
	Asian	12	4.06 (0.86)	25.17 (4.73)	3.51 (1.39)	5.07 (1.04)	92.99 (0.06)	4.42 (1.12)
	European	11	3.47 (0.89)	25.00 (3.97)	3.66 (1.21)	5.14 (1.05)	93.84 (0.08)	4.10 (1.14)
Male	African	14	3.69 (1.03)	27.20 (5.27)	2.90 (1.38)	1.93 (0.92)	93.00 (0.22)	3.55 (1.28)
	Asian	8	3.48 (0.93)	27.25 (5.10)	2.58 (1.25)	2.68 (0.99)	92.44 (0.06)	3.75 (1.12)
	European	11	3.74 (0.86)	26.69 (3.78)	3.76(1.40)	2.26 (1.08)	96.15 (0.05)	3.89 (1.24)

N = number of images. Age is in years, perceived race is percent reporting target item within race category and not multiracial. All other items on a scale of 1–7

are of professional quality, with well-controlled details such as lighting, positioning, hair, and makeup; there are few visible cues to indicate the date and culture from which the images originate, thus preventing rapid dating of the stimuli; they have a naturalistic appearance that is not obviously manipulated or morphed; and they are racially diverse, with multiple instances of both men and women within three different racial categories, and they are more consistent than other recent databases (Ma et al., 2015).

Database development

The faces in the database come from 74 paid volunteers. They were recruited from flyers and in person on two campuses in the Durham area: Duke University and North Carolina Central University. Participants were between the ages of 18 and 25 and either of European, African, or East Asian ancestry (see Table 1).

The models had to meet several requirements in order to be included in the database. They had to have facial features that would unambiguously place them in one of the three racial categories. To minimize the extent to which the images would look specific to time or culture, participants could not have unnatural hair styles or colors, facial hair, or facial piercings. Participants had medium to dark brown eyes, so that there would be no systematic differences in eye color between the races. On the day of the shoot, models were asked to arrive clean-shaven,² without make-up, without any jewelry or hair accessories from the collarbone up (such as necklaces or barrettes), and with their hair worn pulled back and swept completely away from the face and neck. Participants with short hair or bangs were given a light styling gel to allow their hair to be moved off of the face. All subjects wore a collarless black shirt.

²A couple of participants did not follow this instruction. Rather than turn them away, we included them in the database, as this detail may not be important for some researchers.

Before the photographic session, models were given a release form to sign, explaining that their likeness would be used towards the creation of a face database, which would be distributed to other researchers and may appear in academic presentations and publications.

Digital photographs were taken in a professional photography studio by Titus Brooks Heagins, MFA. Photographs were taken with a Canon EOS 1Ds Mark III using a 85 mm F1.2 lens against a plain white paper backdrop. A Sekonic L508 Light Meter was used to ensure optimum and consistent exposure across participants. Studio lighting was created with two Alien Bees B800 Flash Units with umbrellas and one continuous strobe with a soft box.

Photographs were taken in Camera Raw 7.0 (CR2) format resulting in 23.4 × 15.6 in. images with a resolution of 240 pixels per inch. Multiple photographs of each participant were taken. The image selected for inclusion in the database were those that had a good exposure level and focus, a neutral facial expression, the face plane parallel to the angle of camera, the neck and shoulders straight in relation to the head, and minimal hair out of place.

The images were uploaded from the camera's compact flash (CF) card to an iMac desktop computer and edited in Adobe Photoshop CS6. Images were standardized across a variety of dimensions. The faces were centered vertically at the philtrum or septum (whichever lead to greater global bilateral symmetry) and horizontally at the pupils. The size of the face was adjusted so that the height from the bottom edge of the chin to the top edge of the forehead was 2160 pixels (the hairline can sometimes make the starting point of the forehead ambiguous, so this was an approximate guideline). As there is natural variation in the length of the face compared with its width, and in the proportion of the head above and below the eyes, the images vary in how much of the neck is visible, and where there top and sides of the face end. Facial length and pupil location were the only dimensions kept constant between images.



Fig. 1 Sample images from each of the gender/race categories available in the MR2 database. The full-sized images are approximately 15 times larger than displayed in this figure

All faces went through a rigorous digital ‘clean up’ procedure. The faces were rotated to be perfectly horizontal, as determined by the outer corner of the eyes. In some cases, the shoulders were rotated independent of the head, so that the two were straight in relation to each other. Stray hairs were reduced, though not removed completely (to preserve a natural look). Blemishes such as acne, moles, scars, and sores were likewise minimized but not eliminated. Finally, the background color was altered to be a pale neutral gray (RGB 212, 212, 213). See Fig. 1 for sample images from the MR2 database.

The final images are 13 in. (3120 pixels) square, 240 ppi, saved in an uncompressed, lossless TIFF format. The full set is available online at <http://www.ninastrohming.com/MR2>. The MR2 is licensed under a Creative Commons Attribution Non-Commercial ShareALike 3.0 Unported license. The images are free to use, distribute, and alter so long as the following conditions are met. First, the images must be attributed to the authors, citing the present paper.

Second, the images may not be used for commercial purposes. Any alterations or additions to the MR2 must be made available using the same or similar license.

Study 1: stimulus norming

In this study, we gathered norming data on various dimensions of the MR2 that are relevant to conducting psychological research with faces.

Methods

A total of 195 American participants (66 % male, median age 31 years) completed a questionnaire online of which 86 % were of European ancestry, 5 % were of African ancestry, and 4 % were of East Asian ancestry.

Each participant rated 20 randomly selected pictures from the database along six dimensions: estimated age (free response, in years), physical attractiveness (1–7, Not at all attractive to Very attractive), Mood (1 = Negative mood, 7 = Positive mood, 4 = Neutral mood), trustworthiness (1 = Not at all, 7 = Very), Masculinity (1 = Very masculine, 7 = Very feminine, 4 = Neutral). Subjects also categorized the primary race of the person, with an option to include additional races if the person appeared to be of mixed ancestry. Image race and sex were within-participant variables in order to maximize the comparability of norming scores across categories.

Results

Table 1 summarizes the results. There was a high level of agreement that the images fell within the target racial category, while also not being judged as multiracial. No face was primarily judged to belong to a race outside of its target racial category.

Most images had a modal rating of 4 for mood, which is the neutral anchor point ($Mdn = 4$). Since all faces had a target neutral expression, this method of assessing mood is expected to be more sensitive to variance than a within-subject design where faces of other emotional expressions (e.g. sad, happy) are included (Schwarz, 1995).

The subjective ratings show there is variance between the faces when broken down into their respective racial and gender categories. This variance conforms to known differences in how Western populations judge these social groups. For instance, women are considered more trustworthy than men (Todorov et al., 2015), and African Americans are considered more masculine than Asian Americans (Galinsky et al., 2013). Because these differences appear to reflect the natural variance in the judgment of human faces, and because our goal was an



Fig. 2 Stimuli created by layering faces over one another from two face databases, and dividing by race and gender. **a** Images from the CFD. **b** Images from the MR2 used in Study 2. **c** Images from the MR2 not included in Study 2, as there were no Asian faces in the CFD to compare them against

ecologically valid dataset, there was no need to cull the images on the basis of these ratings. The full breakdown of ratings for each item is provided with image download (<http://www.ninastrohming.com/MR2>) for researchers who wish to control for these factors.

Study 2: inter-stimulus consistency

To determine the consistency of the images in our database, we tested it against another recent database with similar properties, the Chicago Face Database (CFD) (Ma et al., 2015). Although the CFD has a slightly lower resolution and does not contain Asian faces, it is a large, contemporary stimulus set.

Consistency refers to inter-stimulus control on dimensions outside of the face itself: similarity in positioning and size, with few extraneous non-facial details. One way to determine the consistency of a set of images is to layer them over one another. To the extent that they are consistent, the resultant layered image will create an

averaged face that is clear and sharp. Inconsistent stimuli will result in an averaged face that is fuzzier and unfocused.

To compare consistency between these two databases, we layered faces from each database by race and gender and asked subjects to compare the pairs of resulting images. We predicted that the MR2 would result in layered images that were judged clearer and less blurry than the CFD, thus reflecting the higher consistency of the MR2.

Methods

Stimulus creation Faces were layered using Photoshop to create a single face for each gender and race combination for each of the two databases. Each new face was added at the occupancy level of $x/1$ (e.g., the third image was layered at 33 % occupancy). We used each image in the MR2, and randomly selected an equal number of images to layer from the CFD (11 European woman, 11 European men, 18 African women, 14 African men; Fig. 2.)

The layered images for the Asian faces in the MR2 are shown in Fig. 2C for the purposes of illustration. Since there were no Asian faces available in the CFD, these images were not tested.

Procedure Two-hundred and one participants completed the study through Amazon Mechanical Turk and 22 failed the informational manipulation check, leaving 179 participants (54 % female; M age = 35; 83 % white, 7 % African American, 6 % Asian, 5 % Hispanic, 1 % Native American; all from United States).

Participants saw four pairs of images (African and European male and female faces), each pair made up of one image from the MR2 and its complement in the CFD. Participants were asked to select which face looked more blurry, fuzzy, and clear (reversed scored) on a scale from 1 (“Definitely Face A”) to 4 (“Definitely Face B”). The three questions were combined to form a single consistency measure, $\alpha > .90$. The order of images within each pair and the order of pairs were randomized between subjects. Higher scores indicate greater blurriness selected for the CFD.

Results

A single one-sample t test with all ratings of the MR2 and the CFD combined resulted in a significant effect, $M = 3.16$, $SD = .54$, $t(178) = 16.31$, $p < .001$, $[.58, .74]$, suggesting greater consistency and less variance in the MR2 relative to the CFD. One-sample t tests comparing coherence ratings of each race gender pair to the scale mean resulted in significant effects for each comparison: European males $M = 3.10$, $SD = .81$, $t(178) = 9.87$, $p < .001$, $[.48, .72]$, European females $M = 3.26$, $SD = .78$, $t(178) = 13.00$, $p < .001$, $[.65, .88]$, African males, $M = 3.00$, $SD = .81$, $t(178) = 8.21$, $p < .001$, $[.38, .62]$, African females, $M = 3.29$, $SD = .76$, $t(178) = 13.88$, $p < .001$, $[.68, .90]$. These results indicate that the MR2 is more coherent than the CF2, both on the whole, and for each of the four race-gender categories.

Discussion

This photo database successfully achieved our stated goals of creating a high-quality, well-controlled, realistic, and racially diverse facial stimulus set with limited context cues. Study 1 confirmed that individual items are judged as similar across a variety of dimensions (e.g., age, mood), while also demonstrating an expected and naturalistic amount of variability between gender and racial divides. Study 2 showed that the MR2 has increased stimulus consistency

compared with another recent database, the CFD, itself already a very highly controlled dataset. The MR2 is thus better controlled at a basic perceptual level than its most similar competitor. These are important considerations for researchers who require exacting control over the features of their facial dataset (e.g., eyes in the same position on the screen, hairstyles that do not become dated or distracting).

In addition to meeting or superseding extant face databases on dimensions that are of broad interest to researchers, the MR2 has two additional virtues that position it as a novel research tool: image resolution and racial diversity. The intricate detail provided in each photograph allows for precise measurement and manipulation of facial features (e.g., pupil size, skin texture) in a way that is simply not possible with other databases. The size of the images also enables researchers to present specific facial features in isolation as stimuli in their own right (e.g., eyes, mouth). Further, we are aware of no other well-controlled databases that contain multiple items representing both sexes within three racial categories. Thus, this database is uniquely suited to study designs that call for direct comparison across three races.

Though our image set was not created to address the following concerns, it does have certain limitations. Subjects maintained a neutral facial expression while being photographed, so these images are not intended for research exploring facial emotion—a key use of facial stimuli in the past (Ekman & Friesen, 1976; Barrett & Bliss-Moreau, 2009). However, since care was taken in ensuring a neutral expression across all images of the corpus (later validated with the norming study), these images are suitable and valid as neutral control faces. As such, they have application to a wide range of uses when emotional expression is not desired.

The volunteers who contributed to the MR2 represent a fairly narrow age window, between 18 and 25. Thus, this database is not suited for those who are specifically interested in aging (Minear & Park, 2004). There is evidence that individuals are most able to remember and discriminate faces that are similar to theirs, an effect which includes age (Mason, 1986; Bäckman, 1991), so these images are not the ideal choice for older research subjects. This database is highly applicable towards standard psychological research, the bulk of which is performed on a college-aged population (Henrich et al., 2010).

A third limitation is that all stimuli are forward-facing and forward-gazing. Some studies manipulate eye-gaze as a way of examining social cognition; the MR2 cannot be used for this purpose without being digitally altered. However, forward-facing/gazing faces are not only maximally true to life—matching how we engage with others—but capture attention most effectively (Senju & Hasegawa, 2005). Forward facing/gazing people are most likely to be seen

as socially relevant, initiating categorization processes most reliably (Macrae et al., 2002).

A final limitation of this database is the total number of images included in the set. While this number (74) is larger than most currently available databases, it is not as large as other sets (Minear & Park, 2004; Ma et al., 2015). Emerging evidence suggests that the power of studies is limited not just by participant sample size but by the size of stimulus sets as well (Westfall et al., *in press*).

Despite these limitations, we believe that this face database provides a valuable resource to psychological science. With its variety of mega-resolution photographs of diverse races, it enables researchers to conduct studies with both experimental and mundane realism. The controlled context of the stimuli ensures that they will remain useful for years to come, and that studies using them will not be confounded by extraneous variables. Faces are arguably the most important source of interpersonal information, and this database presents this information as powerfully and as cleanly as possible.

Author Note The authors would like to thank Crystal Shackelford, Joseph Harvey, Jingxian Zhang, Cynthia Wang, Esiyena Abebe, Irene Lee, and Kara Post for their assistance in recruiting and running subjects.

References

- Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition*, 19(1), 63–71.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “reading the mind in the eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241–251.
- Barrett, L. F., & Bliss-Moreau, E. (2009). She’s emotional, he’s having a bad day: Attributional explanations for emotion stereotypes. *Emotion*, 9(5), 649–658.
- Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, 36(3), 355–370.
- Bellhumer, P. N., Hespanha, J., & Kriegman, D. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7), 711–720.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551–565.
- Chalup, S. K., Hong, K., & Ostwald, M. J. (2010). Simulating pareidolia of faces for architectural image analysis. *International Journal of Computer Information Systems and Industrial Management Applications*, 2(1), 263–278.
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, 13(1), 89–109.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17(5), 383–386.
- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1), 351–362.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Ekman, P., & Matsumoto, D. (1993). *Japanese and Caucasian Facial Expressions of Emotion (JACFEE)*. Palo Alto: Consulting Psychologists Press.
- Emery, N. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6), 581–604.
- Fink, B., Grammer, K., & Thornhill, R. (2001). Human (homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology*, 115(1), 92–99.
- Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, 24(4), 498–506.
- Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56(4), 544–549.
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125, 125–130.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233.
- Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, 3(2), 107–135.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, 33(2-3), 61–83.
- Hugenberg, K., & Wilson, J. P. (2013). Faces are central to social cognition. In Carlson, D. E. (Ed.), *The Oxford handbook of social cognition* (pp. 167–193). Oxford, UK: Oxford University Press.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302–4311.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces - KDEF. CD-ROM.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(1), 1–14.
- Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., & Mason, M. F. (2002). Are you looking at me? Eye gaze and person perception. *Psychological Science*, 13(5), 460–464.
- Mandal, M. K. (1987). Decoding of facial emotions, in terms of expressiveness, by schizophrenics and depressives. *Psychiatry: Interpersonal and Biological Processes*.
- Mandal, M. K., Harizuka, S., Bhushan, B., & Mishra, R. (2001). Cultural variation in hemifacial asymmetry of emotion expressions. *British Journal of Social Psychology*, 40(3), 385–398.
- Mason, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research*, 12(3), 151–154.
- Milborrow, S., Morkel, J., & Nicolls, F. (2010). The MUCT landmarked face database. *Pattern Recognition Association of South Africa*. <http://www.milbo.org/muct>
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4), 630–633.
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35.

- Nordström, M. M., Larsen, M., Sierakowski, J., & Stegmann, M. B. (2004). The IMM face database: An annotated dataset of 240 face images. Technical report, DTU Informatics, Building 321.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ..., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*(1), 36–88.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092.
- Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75–92.
- Perrett, D., May, K., & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature, 368*(6468), 239–242.
- Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20* (5), 295–307.
- Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing, 16*(5), 295–306.
- Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition, 20*(2), 143–169.
- Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception, 32*(9), 1093–1107.
- Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: facial maturity modulates recognition of fearful and angry expressions. *Emotion, 9*(1), 39.
- Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review, 63*(2), 153–177.
- Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition, 12*(1), 127–144.
- Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics, 72*(3), 695–705.
- Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452–460.
- Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior, 27*(2), 741–749.
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*(1), 519–545.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460.
- Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution, 52*(3), 314–320.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249.
- Wang, L., & Markham, R. (1999). The development of a series of photographs of Chinese facial expressions of emotion. *Journal of Cross-Cultural Psychology, 30*(4), 397–410.
- Westfall, J., Judd, C. M., & Kenny, D. A. (in press). Replicating studies in which samples of participants respond to samples of stimuli. Perspectives on Psychological Science.
- Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: Real but not artificial faces sustain neural activity beyond the n170/vpp. *PLoS ONE, 6*(3), e17960.
- Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3D facial expression database for facial behavior research. In *7th International Conference on Automatic Face and Gesture Recognition* (pp. 211–216): IEEE.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.
- Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception-London, 31*(8), 995–1004.