

A new theoretical approach to improving face recognition in disorders of central vision: Face caricaturing

Research School of Psychology, Australian National University,
Canberra, Australian Capital Territory, Australia
ARC Centre of Excellence in Cognition and Its Disorders,
Australian National University, Canberra,
Australian Capital Territory, Australia

Jessica Irons



Research School of Psychology, Australian National University,
Canberra, Australian Capital Territory, Australia
ARC Centre of Excellence in Cognition and Its Disorders,
Australian National University, Canberra,
Australian Capital Territory, Australia

Elinor McKone



Rachael Dumbleton

Research School of Psychology, Australian National University,
Canberra, Australian Capital Territory, Australia



National Information and Communication Technology
Australia (NICTA), Canberra,
Australian Capital Territory, Australia
College of Engineering and Computer Science, Australian
National University, Canberra,
Australian Capital Territory, Australia
Bionic Vision Australia, Carlton, Victoria, Australia

Nick Barnes



National Information and Communication Technology
Australia (NICTA), Canberra,
Australian Capital Territory, Australia
College of Engineering and Computer Science, Australian
National University, Canberra,
Australian Capital Territory, Australia

Xuming He



John Curtin School of Medical Research and Medical School,
Australian National University, Canberra,
Australian Capital Territory, Australia

Jan Provis



Callin Ivanovici

Research School of Psychology, Australian National University,
Canberra, Australian Capital Territory, Australia



Alisa Kwa

Research School of Psychology, Australian National University,
Canberra, Australian Capital Territory, Australia



Citation: Irons, J., McKone, E., Dumbleton, R., Barnes, N., He, X., Provis, J., Ivanovici, C., & Kwa, A. (2014). A new theoretical approach to improving face recognition in disorders of central vision: Face caricaturing. *Journal of Vision*, 14(2):12, 1–29, <http://www.journalofvision.org/content/14/2/12>, doi: 10.1167/14.2.12.

Damage to central vision, of which age-related macular degeneration (AMD) is the most common cause, leaves patients with only blurred peripheral vision. Previous approaches to improving face recognition in AMD have employed image manipulations designed to enhance early-stage visual processing (e.g., magnification, increased HSF contrast). Here, we argue that further improvement may be possible by targeting known properties of mid- and/or high-level face processing. We enhance identity-related shape information in the face by caricaturing each individual away from an average face. We simulate early- through late-stage AMD-blur by filtering spatial frequencies to mimic the amount of blurring perceived at approximately 10° through 30° into the periphery (assuming a face seen premagnified on a tablet computer). We report caricature advantages for all blur levels, for face viewpoints from front view to semiprofile, and in tasks involving perceiving differences in facial identity between pairs of people, remembering previously learned faces, and rejecting new faces as unknown. Results provide a proof of concept that caricaturing may assist in improving face recognition in AMD and other disorders of central vision.

Introduction

This article presents a new general approach that may be useful in improving face recognition in people with permanently blurred vision. The most frequent cause of such damage is age-related macular degeneration (AMD). AMD is the leading cause of vision impairment in many developed countries (e.g., one in seven over the age of 50 in Australia; Deloitte Access

Economics, 2011), with a global healthcare cost in 2010 of \$255 billion USD (Access Economics, 2010). AMD causes deterioration of the macula, the central area of the retina (for reviews see de Jong, 2006; Lim, Mitchell, Seddon, Holz, & Wong, 2012). As a result, patients with AMD often experience a scotoma in central vision. This requires them to rely on low resolution peripheral vision, with reduced visual acuity and reduced contrast sensitivity particularly at high spatial frequencies (e.g., Kleiner, Enger, Alexander & Fine, 1988; Sjöstrand & Frisén, 1977). Moreover, vision typically becomes more blurred with disease progression as the damaged region of the visual field extends out to greater eccentricities (Sunnus et al., 1999). Later-stage AMD scotomas commonly exceed 20° diameter (Cheung & Legge, 2005).

In addition to impairing everyday tasks such as reading and driving, AMD impairs the ability to tell apart individual people from their faces. In two self-report studies, 40% of AMD patients reported at least some difficulty in recognizing the facial features of a person at arm's length, 75% reported difficulty for people across the other side of a room, and 97% reported difficulty recognizing people on the street (Schmier & Halpern, 2006; Tejeria, Harper, Artes, & Dickinson, 2002). In an objective test, participants with AMD required faces to be substantially larger than controls to compensate for their low-resolution vision (Bullimore, Bailey, & Wacker, 1991).

A simulation by Marmor and Marmor (2010), shown in Figure 1, illustrates that face recognition deficits present in AMD are as would be expected from the retinal damage. Reliably identifying individual faces—for example, distinguishing between different

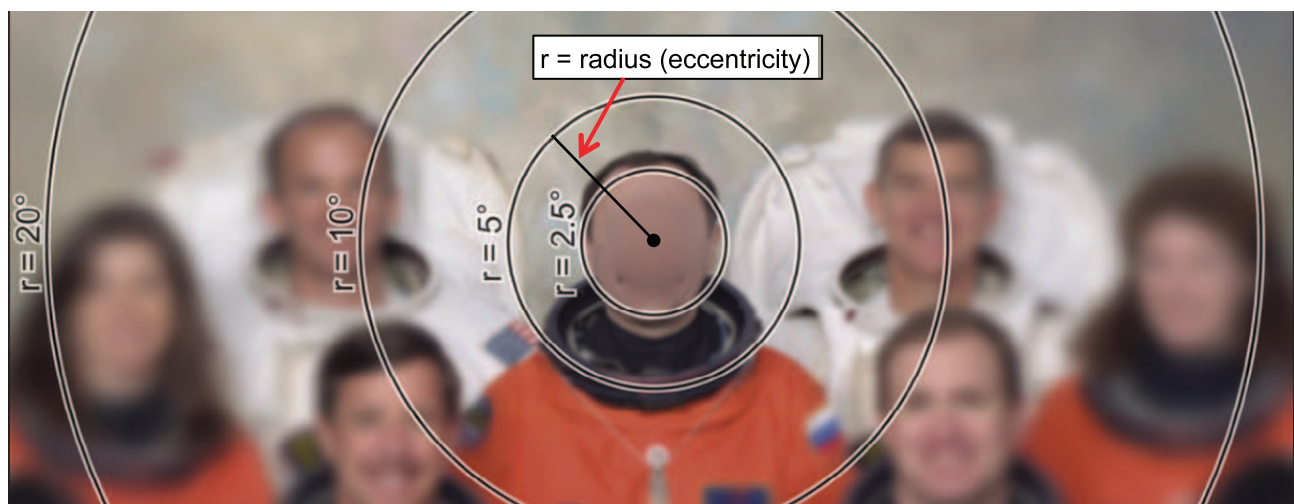


Figure 1. Marmor and Marmor's (2010) simulation of the increased blur present in faces with increasing eccentricity, illustrating the corresponding difficulty in recognizing facial identity where a patient has a central scotoma. The astronauts are assumed to be 2.7 m away from the viewer. Adapted with permission from Marmor and Marmor (2010). Copyright © 2010 American Medical Association. All rights reserved.

Caucasian young adult men all with short brown hair—is very poor based on low spatial frequency information alone (<5 cycles per face), and to achieve the best performance requires both medium (5–8 cycles per face), and high (8–12 cycles per face) spatial frequency information (Fiorentini, Maffei, & Sandini, 1983). High spatial frequency processing is best in central vision and becomes progressively poorer with increasing eccentricity in peripheral vision. Thus, a central scotoma will impair face recognition because faces viewed in the periphery, at a fixed size, will appear increasingly blurred at greater eccentricities (Figure 1). Moreover, as AMD disease progression increases the radius of the scotoma, the face recognition difficulties will become more severe.

Losing the ability to recognize faces can have an important negative impact on social interaction. Research into prosopagnosia—impaired facial recognition due to atypical functioning of the cortical face recognition system—shows that inability to recognize faces can cause stress and anxiety in social situations, leading to social avoidance and isolation (Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008). Even within the normal population, poorer performance on face recognition tasks is correlated with poorer social skills and higher social anxiety (Davis et al., 2011; Li et al., 2010). The negative effects of poor face recognition are likely to be especially significant in AMD, considering that people with AMD are a psychologically vulnerable group (e.g., increased rates of depression; Brody et al., 2001).

Theoretical approaches to improving face recognition in AMD

Despite the everyday importance of face recognition, to date there has been little consideration of whether face recognition in people with AMD or other disorders of central vision could be improved. To our knowledge, only two methods have been explored.

The first is magnification; that is, making the face image larger to reduce the amount of perceptual blur. Tejeria et al. (2002) used telescopes attached to the top of the wearer's glasses to allow switching between normal viewing through the glasses and magnified viewing through the telescope. AMD performance in naming celebrities improved with the telescopes. However, vision through telescopic devices can be difficult to adapt to, and many low vision patients have been hesitant to adopt them (Lowe & Rubinstein, 2000; Tejeria et al., 2002). The second approach is increasing the contrast of the medium and high spatial frequency information in the face images. This approach (Peli, Goldstein, Trempe, & Arend, 1989; Peli, Goldstein,

Young, Trempe, & Buzney, 1991) has shown that amplifying higher spatial frequencies led to approximately half of AMD patients showing significant gains in recognizing famous faces. Some participants also reported that the enhanced images appeared clearer and easier to see.

From a theoretical perspective, both of these previous techniques—magnification and increased higher-spatial-frequency contrast—are directed towards enhancing the representation of faces in *early* visual processing areas (e.g., V1). This general approach uses the logic that enhanced low-level coding will then flow through to provide higher quality input to later visual areas, including the high-level areas in which individual faces are eventually recognized.

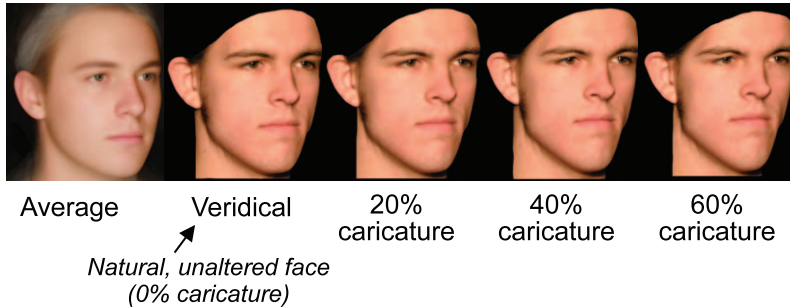
Importantly, however, logically it is equally possible to try to improve face recognition by altering stimuli to best match the properties of face coding in *mid-* and *high-level* visual processing areas. There are many mid- and high-level areas that are responsive to faces and are likely to contribute to overall recognition performance (e.g., V4, lateral occipital complex, fusiform face area, occipital face area; for review see Kanwisher & Dilks, 2013). Neural responses in these areas are typically much less sensitive to changes in stimulus size or stimulus contrast than coding in early visual areas.

Theoretically, we know a great deal about higher-level perceptual coding of faces, derived primarily from behavioral studies. This extensive literature has, for example, argued that faces are coded in several important ways, including holistically (e.g., Rossion, 2013; Young, Hellawell, & Hay, 1987), additionally as local parts (deGutis, Wilmer, Mercado, & Cohan, 2013; Pitcher, Walsh, Yovel, & Duchaine, 2007), and also as deviations from an average in a perceptual face-space (e.g., Susilo, McKone, & Edwards, 2010; Valentine, 1991). Potentially, stimulus alterations targeted to any of these higher-level coding styles might be able to improve face recognition in AMD. Here, we begin the process of exploring such approaches, by examining one type of stimulus alteration—caricaturing—targeted at improving one type of higher-level face representation, face-space coding.

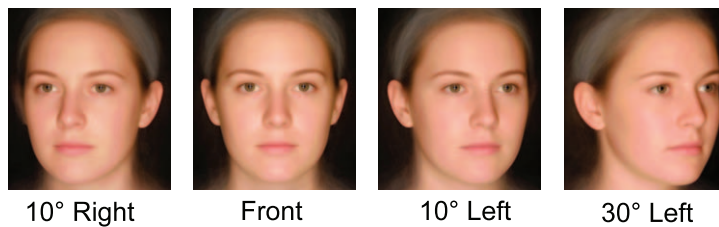
A high level approach to face stimulus enhancement: Caricaturing

In the present study, we test a combination of magnification (via electronic rather than telescopic means) together with image enhancement via caricaturing. *Caricaturing* (Figure 2) is a process whereby the ways in which an individual's facial shape differs from an average face are exaggerated. For example, if a face is slightly narrower, has a slightly more pointed nose,

A. Caricaturing a natural face away from the average



B. Multiple averages so only identity is caricatured.



C. Face-space explanation of why caricaturing improves face recognition

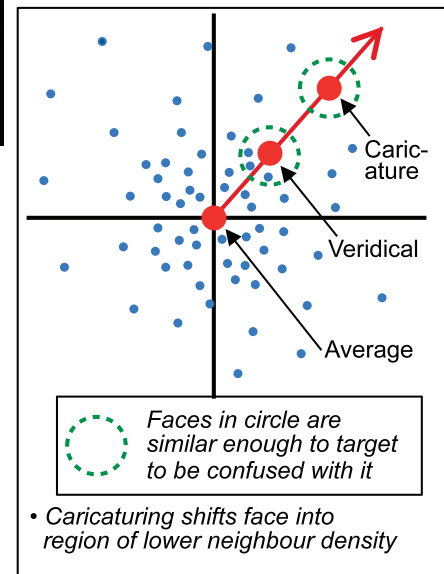


Figure 2. Caricaturing and face-space. (A) The process of making a face caricature. The veridical face is morphed away from an average face (average of many individuals), such that all aspects of the face are exaggerated. In this individual, such aspects include the long chin, the tilted tip of nose, the straight jaw, the closeness of eyebrows to eyes, and so on. Note that only shape, not color, is caricatured in our stimuli. (B) To ensure that only face identity information was caricatured, all our faces had neutral expression and were one race (Caucasian), and we used separate averages for each viewpoint for males (A; see Figure 11 for male averages in all viewpoints) and females (B), with eight average faces total. (C) Face-space explanation of where caricatured faces lie in face-space, and why this leads to improved ability to recognize the face. Blue dots indicate individual faces, coded in terms of their value relative to the average on multiple face attributes (Note: It is unknown what the specific dimensions are, and only two are illustrated here). Caricaturing shifts the face into a region of lower exemplar density, meaning that there are fewer confusable neighbors.

and eyes slightly closer together than does the average face, then the caricature of that face will become narrower still, have an even more pointed nose, and eyes even closer together. Caricaturing targets mid- and/or high-level cortical vision, by enhancing the *shape* information in a stimulus. Prior research with high-resolution images has shown that caricatured faces are commonly easier to recognize than the unaltered (veridical) face.

Face caricatures for high-resolution photographs are created using morphing software (see Benson & Perrett, 1991). This process involves comparing a target face photograph with an average face (an image created by morphing together numerous faces sharing the same viewpoint, expression, sex, age, and race as the target; Figure 2A & B). Multiple key locations are marked on both faces with corresponding points. The distance between each point on the target face image and the corresponding point on the average face image is then increased. The result is that the entire facial shape of the target is morphed in a direction away from the average and becomes more exaggerated. Because the average face is matched to the target face on factors such as age and sex, it is only the shape attributes that

are idiosyncratic to the target's *identity* that become exaggerated. Various levels of caricature can be created, depending on the degree to which the distances between points are increased (see Figure 2A).

Experimental studies of young adults tested with high-resolution (i.e., unblurred) images have established that caricatured photographs of faces are commonly better recognized than the veridical (unaltered) version (Benson & Perrett, 1991; Calder, Young, Benson, & Perrett, 1996; Chang, Levine, & Benson, 2002; Lee, Byatt, & Rhodes, 2000; Lee & Perrett, 2000; Rhodes, Byatt, Tremewan, & Kennedy, 1997). For example, Lee et al. (2000) found caricatured photographs of male celebrities were more accurately identified than veridical images. A similar caricature advantage has been demonstrated on tasks in which participants learn previously unseen faces and are then tested on their recognition of these faces (Chang et al., 2002; Rhodes et al., 1997).

Theoretically, the caricature advantage is commonly understood within a perceptual face-space (see Figure 2C; Valentine, 1991), supported by both psychophysical (e.g., Blank & Yovel, 2011; Susilo et al., 2010; Valentine & Bruce, 1986) and neuroscientific (e.g.,

Freiwald, Tsao, & Livingstone, 2009; Leopold, Bondar, & Giese, 2006) evidence. According to this framework, individual faces are coded as points in a multidimensional perceptual space, the dimensions of which represent facial attributes tuned to the set of previous faces to which the observer has been exposed. The physically average face lies at the center of the space. The space contains a higher density of face exemplars (people previously learned from everyday life) closer to the center of face-space (i.e., faces with typical values on most dimensions) than further away (i.e., faces with atypical values) (Johnston, Milne, Williams, & Hosie, 1997). When a face is caricatured, it shifts along its original vector further away from the average face as compared to the veridical face (Figure 2C), and thus moves into a region of lower exemplar density. Valentine (1991) argued that lower exemplar density leads to a face being easier to recognize, because it will have fewer nearby neighboring exemplars with which it can be confused.

Although it is well established that caricaturing can improve recognition of high-resolution images, its effect on blurred images similar to those experienced by AMD patients has not previously been examined. Encouragingly, a number of studies have explored caricature effects using face images that are visually degraded in some way, and in general these have found the caricature advantage to still be present. This includes for line drawings (Rhodes, Brennan, & Carey, 1987; Rhodes et al., 1997; Stevenage, 1995), and for face stimuli of very brief duration (33 ms followed by mask; Lee & Perrett, 2000). However, neither of these manipulations is logically similar to that experienced in AMD. Line drawings retain high spatial frequencies, while AMD specifically impairs high spatial frequencies. Moreover, AMD patients do not see faces only briefly, but rather have as long to observe and scan them as normal controls.

Simulating the peripheral blur of AMD

The goal of the present study was to provide a proof of concept examining whether caricaturing might be an effective technique for improving face recognition in people with AMD. Importantly, we tested a *simulation* of central vision impairment, focusing on its key aspect, namely perceptual blur. Specifically, we tested young adult observers with normal vision, shown blurred images in central vision that were designed to mimic the appearance of faces at various retinal eccentricities. The levels of blur we tested, which were created using spatial filtering to progressively remove spatial fre-

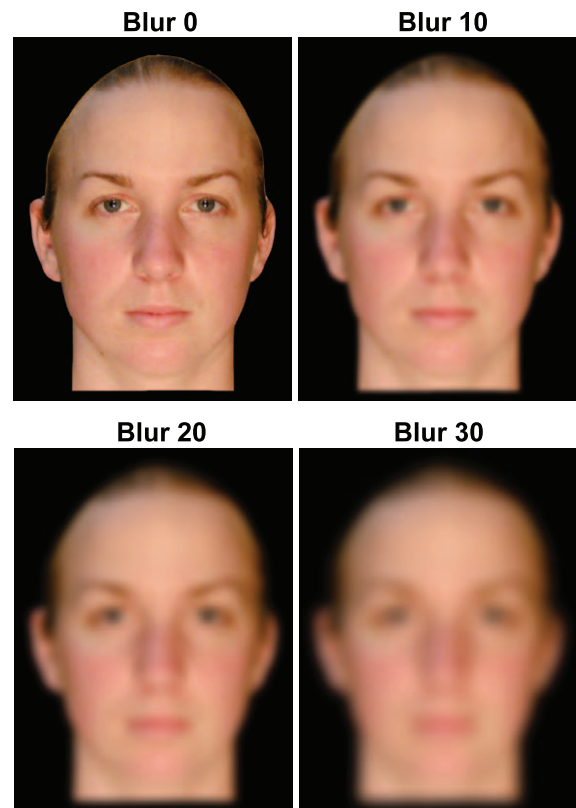


Figure 3. The levels of blur used in the present study, designed to simulate the degree of blur present when viewing a face (18 cm wide ear-to-ear and at 40 cm distance, which is equivalent to a real person seen 54 cm away in the real world) at 0°, 10°, 20°, and 30° eccentricity (Blur0, Blur10, Blur20, Blur30, respectively).

quencies starting from the highest, are illustrated in Figure 3.

Importantly, we note that adding blur does not capture all aspects of the way AMD patients see faces. Blur is the most obvious difference between normal central vision and AMD (peripheral) vision, but there are also other important differences. In general, however, these are difficult to simulate in normal observers. For AMD patients who develop a preferred retinal location (PRL) for looking at the visual world, patterns of their “fixations” to a face stimulus differ from the patterns of foveal fixations in normal-vision observers (Seiple, Rosen, & Garcia, 2013). At first glance, it might seem possible to simulate this aspect of AMD, for example by tracking a normal-vision observer’s eye movements and using this information to mask the region of the stimulus falling on the fovea, thus forcing the normal observer to use peripheral vision to scan the faces. In fact, however, such an approach would not provide an appropriate simulation of face processing in AMD patients. In AMD, extensive changes in neural response occur such that large areas of early visual cortex that are normally

responsive only to input from central vision (in the foveal confluence near the occipital pole) instead become responsive to peripheral stimuli, possibly because of the “unmasking” of previously inhibited connections (e.g., Baker, Dilks, Peli, & Kanwisher, 2008; Dilks, Baker, Peli, & Kanwisher, 2009; Schumacher et al., 2008). Because normal vision observers lack these changes, the cortical processing of faces presented to the periphery in normal vision observers is very different from the cortical processing of faces viewed with the periphery by AMD patients. Also, presenting faces peripherally to normal observers would be unlikely to replicate the eye movement patterns of AMD because normal vision observers would not have a PRL (e.g., in half of AMD patients developing a PRL, it takes more than 1 month of practice using peripheral vision; Crossland, Culham, Kabanarou, & Rubin, 2005).

Overall, we argue that although adding blur to face images necessarily provides a less than perfect simulation of AMD, it captures the core feature of the vision problem in AMD. Also note that simulations (again, less than perfect) are well accepted as a valuable method of research in other areas of low vision (e.g., prosthetic vision; Dagnelie, 2008; van Rheede, Kennard, & Hicks, 2010). Simulations offer practical advantages in being able to explore more stimulus and learning situations than is possible when testing real patients. This can then provide a basis for maximizing efficiency and ethics of testing the successful techniques in patients.

Present study

Our experiments are structured as follows. First, we assess the caricature advantage in face *perception*, measured as the increase in perceived dissimilarity between pairs of identities (Experiment 1). Then, we shift to a direct test of individuation performance in face *memory* (improvement in old–new recognition). We assess memory under two different learning regimes (subject learns each person *either* as caricature *or* veridical [see Experiment 2] and learns each person as *both* caricature *and* veridical [Experiment 3]). The question of interest in all cases was the amount by which caricaturing enhances face individuation at the different levels of blur, with primary interest in the blurred-face conditions (i.e., those relevant to AMD).

Several aspects of our design and analyses were driven specifically by our practical interest in AMD. First, the specific eccentricities (10° through 30°) we simulated were chosen to match the residual function of the various stages of severity in disease progression of AMD. Second, we presumed that, within the next few

years of developments in computer science, that it might be possible to caricature faces in real time so that patients would be able to view them on a tablet computer held in the crook of the arm; this would mean that patients would view the faces sized at 18° of visual angle (ear-to-ear; see Method for details). Thus, for the blurring procedure we used a face size of 18° at each eccentricity, noting that faces appear more blurred or less blurred at a given eccentricity depending on how large they are, and therefore that simulating the level of perceptual blur at a given eccentricity also requires defining a face size (Marmor & Marmor, 2010). Third, the interest in testing the learn each person *both* veridical *and* caricatured regime (Experiment 3), which is not a standard procedure used in the caricaturing literature, derived from an interest in eventually allowing AMD patients to adjust caricature levels using a slider bar on the tablet computer, meaning that the patient would see each person’s face veridical as well as caricatured.

Experiment 1: Perception of facial identity differences

Experiment 1 used a rating method in which participants were presented with faces of two people simultaneously and asked to rate how similar or dissimilar the two people appeared (Figure 4). To the extent that caricaturing improves individuation, the two people will be perceived as more dissimilar when they are both caricatured than when they are both veridical (the natural unaltered face). This is because caricaturing shifts each face further from the average along its own trajectory, which shifts the two faces further from each other; faces that are further apart in face-space are rated as more dissimilar (Johnston et al., 1997; Lee et al., 2000). Increased dissimilarity ratings with caricaturing have previously been confirmed for high resolution (unblurred) faces (Lee et al., 2000).

Here, we selected the rating method due to its practical benefits. Similarity ratings are, compared to memory scores, generally very stable and show little variance (e.g., see Light, Kayra-Stuart, & Hollander, 1979, who used 14 participants for pairwise similarity ratings, but 30 per condition for memory tasks). This means that, with only a small number of participants, tested for only 1 hr each, we were able to gain neat and reliable data across a total of 16 conditions: four levels of caricature ranging from veridical (0%) to highly caricatured (60%); and four levels of blur ranging from none (i.e., normal central vision) to 30° eccentricity.

For unblurred images, we expected to find that increasing caricature level would be associated with an increase in perceived dissimilarity (Lee et al., 2000).

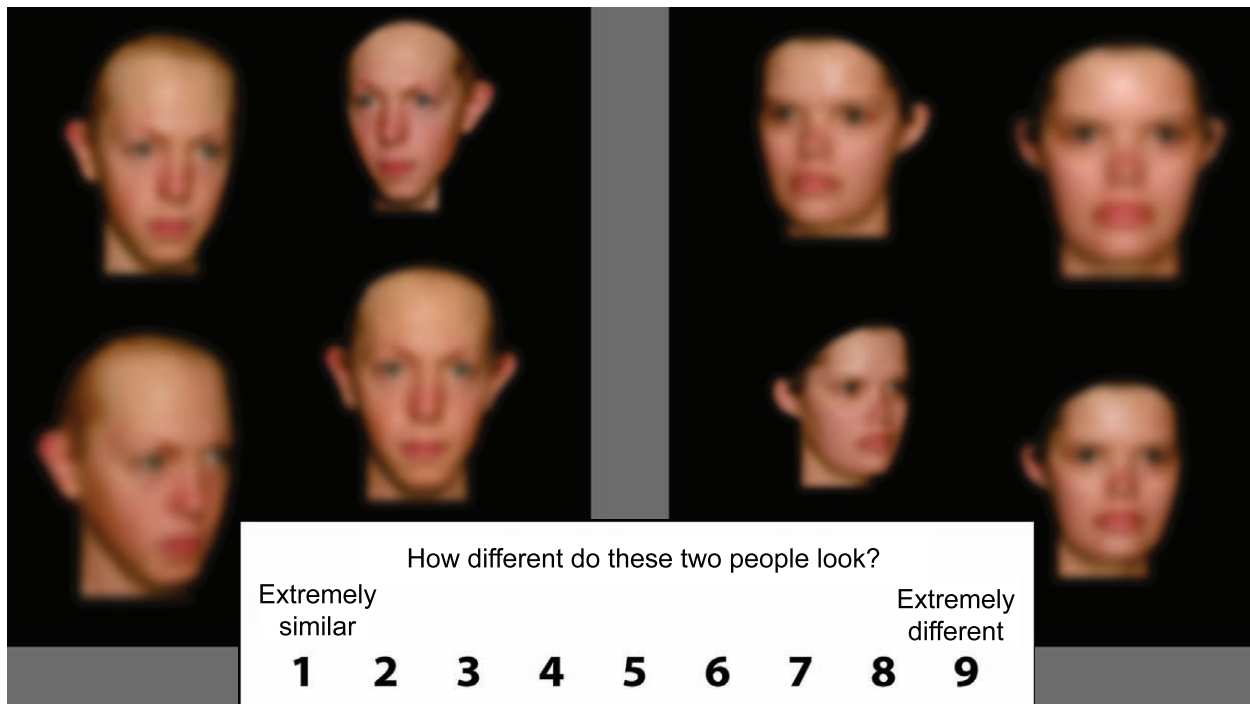


Figure 4. Rating task method for Experiment 1, illustrated using faces in the 20% caricature Blur30 condition.

Our questions of interest were then (a) whether this caricature advantage also occurred for each level of blurring in faces, (b) whether the amount of caricature advantage changed with the amount of blur (i.e., strengthened, weakened, or remained constant), and (c) for each level of blur, whether there was any caricature level for which the benefit was strong enough that it returned face individuation to “normal” levels, namely to the level of dissimilarity perceived in unblurred veridical faces.

Method

Participants

Data reported are from 12 young adult Caucasian students from the Australian National University (nine female and three male; age range 18 to 24 years, $M = 19.92$, $SD = 2.02$) who received psychology first-year course credit or \$15 AUD for the 1 hr experiment.¹ All reported normal or corrected-to-normal vision. Participants were tested wearing their usual correction where applicable; under these circumstances visual acuity, as measured by ETDRS eye charts positioned 10 feet from the viewer, was higher than 20/32 in both eyes for all participants (ranging from 20/12 to 20/32, $M = 18.33$, $SD = 4.25$). The research methods of all experiments adhered to the Declaration of Helsinki.

Design

Each participant was tested on four levels of caricature (0% or veridical, 20%, 40%, and 60%), crossed with four levels of blur (Blur0, Blur10, Blur20, and Blur30, designed to mimic appearance at 0°, 10°, 20°, and 30° eccentricity), in a fully repeated measures design. To encourage similarity judgments based on the perception of the *face*, and not merely the particular photograph of that face, each trial presented four images of each person, shown in four different viewpoints (front view, 10° rotation left from front, 10° rotation right from front, and 30° rotation left from front). The task was to compare two people presented simultaneously in this format (always with both people at the same caricature and blur level; Figure 4) and rate how different the two people looked on the scale of 1 (extremely similar) to 9 (extremely different). A higher score indicates the people were perceived as more dissimilar.

Stimuli

The veridical face images were photographs of Caucasian young adults, selected from the ANU face database (this database contains individuals photographed in Canberra, thus ensuring the within-Caucasian ethnicity of the faces was well matched to our Canberra participant sample; see McKone et al., 2012 for discussion). All faces had neutral expression, closed

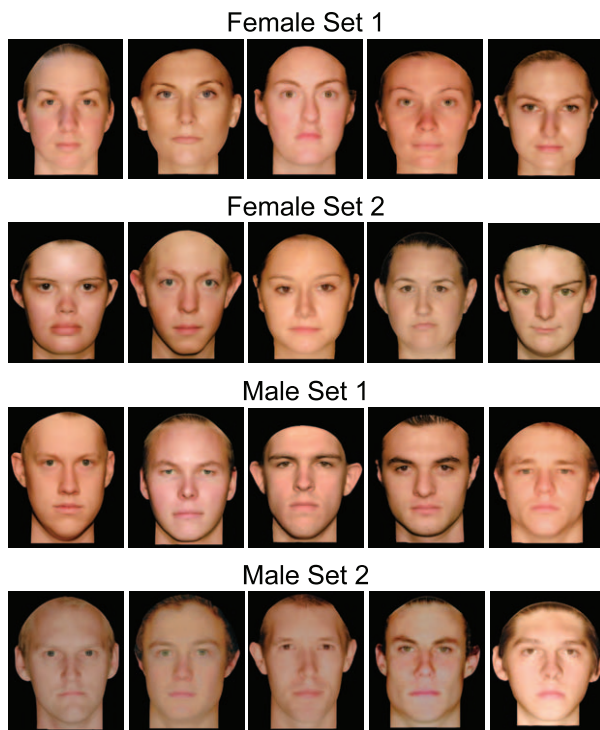


Figure 5. The 20 faces used in our experiments. In Experiment 2 (ratings), ratings were conducted within each set shown (i.e., within Set 1, each woman was rated for similarity to each other woman in turn). In the memory tasks (Experiments 2 and 3), Female Set 1 and Male Set 1 were used as old faces, and Female Set 2 and Male Set 2 as new faces.

mouth, and were without facial hair, obvious makeup, or other salient characteristics (such as piercings) that might influence similarity ratings but are not face information per se. Tilt adjustments were made if necessary so that all heads were fully upright. All images were in color.

Average faces

The first step towards making the face caricatures was to create average faces. Eight separate average faces were created: one for each of the four viewpoints (front, 10° left, 10° right, 30° left) separately for males and females (Figure 2B for female averages, Figure 2A for 30° left male average). A total of 57 female and 26 male individuals were used to make the averages (there were fewer males merely because the database contained fewer male images; note that even 16 faces is sufficient to make a reasonably reliable average, after which adding more individuals from the same ethnicity/sex/age group produces only minor change in the average; Langlois & Roggman, 1990). For most of these, images at each of the four different views were included in the final average; however, for 12 women

and 2 men, one or more of the views were excluded from the averages for various reasons (e.g., the image was blurry, mouth was slightly open).

Average faces were then created using Abrosoft FantaMorph 5 (Abrosoft Co., Beijing, China). Key points were placed by hand on each face, and these points were aligned across all the different images in each sex and viewpoint. The number of points assigned varied between 155 and 175, depending on sex and viewpoint. The faces were then morphed together to create the eight averages.

Caricatures

For the rating task, 20 individuals from the ANU face database were used, 10 males and 10 females (Figure 5), who were a subset of those used to make the averages. The 20 individuals were selected according to the following criteria: (a) photographs at all four viewing angles were of good quality (e.g., not blurry, eyes fully open, no teeth showing); (b) all images closely matched the viewing angle of the corresponding average face (to avoid caricaturing any slight differences in viewing angle), and (c) the lighting and colors were approximately equal within a set when separated into four face sets (two sets of five female faces, two sets of five male faces) so that relatively similar looking faces could be placed together in the same set. The images of each of the 20 targets at the four viewpoints, plus the eight average faces, were placed on a uniformly sized black background. The size, position, and rotation of each target face were adjusted slightly so that the pupil location and interocular distance matched that of the corresponding average face.

To create caricatures, a target image and the corresponding average face were uploaded into FantaMorph (Abrosoft Co.). Key points were placed on the face by hand, ensuring that each point appeared in the same location on both the target and average image, and also across different viewpoints and different individuals (i.e., a point placed on the tip of the nose always appeared on the tip of the nose at all viewpoints and for all individuals; note that exactly matching the location of the key points across viewpoints was necessary to ensure FantaMorph (Abrosoft Co.) produced equivalent caricatures in all viewpoints, e.g., so that the 60% caricature of Face A at 30° left looked like a valid three-dimensional head rotation of the 60% caricature of Face A at 10° right). Nonfrontal faces (10° right, 10° left, 30° left) used 136 points and front-on faces used 147 points (front-on faces included points around both ears, while nonfrontal faces included only the ear closest to the camera). If a caricatured image produced morphing artifacts (e.g., jagged shape boundaries, unexpected straight lines across the image), extra points were added on a case-by-case basis until

the morphing artifacts disappeared. The shape information in the target images was then morphed away from the average using the FantaMorph (Abrosoft Co.) track curve function, which computes and exaggerates the differences between the locations of the key points on the two images. The value of the caricatures was based on the extent to which these differences were increased, where a 0% caricature represented the veridical image and a 100% caricature represented an image in which the differences were doubled. Note that only shape information was caricatured; there was no caricaturing of color information, with all caricatures assigned the color information of the veridical face.

Along with the veridical image, three caricature levels were saved for each image: 20%, 40%, and 60%.² In a handful of cases, the caricatured faces were of noticeably different overall size from the veridical; we adjusted these by hand to their original size (given that absolute face size is not a reliable cue to identity because it varies with distance from the person). All stimuli were then cropped around the hairline and ears, with a straight horizontal crop under the chin, and placed on a black background. Hairstyle and clothing were removed because these are changeable aspects of people's appearance that thus do not form reliable cues to their identity. Removal ensured that any caricature effect reflected perception specifically of the face.

The above method produced 320 images (20 individuals \times 4 viewpoints \times 4 caricature levels), each at 1000 \times 1200 pixels resolution.

Spatial filtering (blurring) to mimic facial appearance in AMD

To produce the final experimental images (1,280 total), each of the 320 was then created in four blur levels. Blur0 had no manipulation applied; that is, the images retained their full spatial frequency content present in the original image. Blur10, Blur20, and Blur30 (illustrated in Figure 3) were designed to mimic the approximate appearance of faces seen at 10, 20, and 30 degrees into the periphery, and thus the residual vision present in AMD with varying stages of disease progression.

The amount of blurring with eccentricity is defined in cycles per face. To implement the blurring kernel in pixel space, we use the size of the face stimulus and the distance to display to calculate the blurring kernel width in pixels. Our 10°, 20°, and 30° eccentricity levels are as for a 12.75 cm face stimulus (across the widest part of the nonbackground region of each image separately, i.e., ear-to-ear in front-view faces) viewed from a distance of 40 cm. This corresponds to a face subtending 18.11° along the horizontal (equivalent to a real person viewed at 54 cm; McKone, 2009), which is a substantial magnification of the face relative to

most natural viewing conditions in everyday life, and results in less blurring than for smaller faces (e.g., our 20° images in Figure 3 are noticeably less blurred than are the astronauts in Figure 1 at 20°, which is because the astronauts have been blurred as they would appear when 2.7 m away, i.e., each head approximately 5° tall). This 18.11° size was based on the rationale that if future software developments allow automatic high quality face caricaturing (see General discussion for current limitations), then a plausible practical scenario could be real-time photographing and caricaturing of real-world faces on a tablet computer. We took an iPad (Apple, Cupertino, CA) of screen size 20 cm tall \times 15 cm wide, and we assumed the face width to be 85% of the screen width (i.e., 12.75 cm) so that natural differences in face aspect ratio between different individuals meant that no normal face would exceed the screen in height. The 40 cm viewing distance is based on the assumption that a user would hold the iPad in the crook of his or her arm.

Image blurring was applied by reducing the contrast of spatial frequencies higher than a specified threshold. This threshold became progressively lower as the simulated eccentricity increased. In AMD, no visual information comes from the region of the scotoma. If we assume that the central fovea is lost, then the individual will be unable to use the area of highest photoreceptor density for vision, and thus will have reduced acuity. Cell density in the retina, and corresponding visual acuity, reduces with distance from the central fovea, or increasing eccentricity. Let us define that the closest functional part of the retina to the fovea is at eccentricity e . The highest acuity that can then be achieved is affected by the projection of cell density, image quality, and the degree of convergence to the receptors on latter processing units. This acuity can be defined by a cutoff frequency, the highest frequency that can be recovered as a visual angle (in cycles per degree). A significant number of papers have plotted visual acuity against retinal eccentricity based on human data including Wertheim (1891/1980), Mandelbaum and Sloan (1947), Millodot (1966), Anstis (1974), and Anderson and Thibos (1999), with some variations in the final numbers based around what is measured and individual performance. Marmor and Marmor (2010) take the results from these studies that mostly concern younger individuals and take a lower bound across them.

In this paper, we simulate the acuity–eccentricity relation by removing the frequency components higher than the cutoff frequency at an eccentricity e . We implement this by applying a uniform spatial blur across the image using a Gaussian kernel filter of size defined by the cutoff frequency. It is well known that the cutoff frequency f (cpd) follows an inverse law with respect to the eccentricity (Anstis, 1974; Peli, Yang, &

Goldstein, 1991; Rovamo, Virsu, & Näsänen, 1978). We represent this as $f = f_0 / (1 + \beta e)$, and set the parameters as $\beta = 0.645$ and $f_0 = 30$ in this study. The resulting curve provides a close fit to the values presented in Marmor and Marmor (2010).

Achieving theoretically ideal frequency cutoff using image filtering is difficult to implement without introducing artifacts; we adopt the conventional Gaussian filters with a kernel width σ_f as $(1/3)f$. This ensures that almost all the frequency components beyond f will be removed. Given the value of σ_f and the physical parameters used in the study, we convert the kernel width to pixel unit as follows. Let the width of screen be w cm, the distance between viewer and screen be d cm, and the horizontal resolution of the image be r pixels. Based on Fourier transform theory, the Gaussian kernel width in the spatial domain is $\sigma_s = 1 / (2\pi\sigma_f)$ degrees. Therefore, we can compute the filter kernel width as $\sigma_s = 3(1 + \beta e) / 2\pi f_0 \times r / 2 \arctan(w/2d)$ pixel.

Our values of 10°, 20°, and 30° eccentricity are only approximate as a simulation of the degree of blur in faces perceived by AMD patients. This is because: (a) the precise formula relating eccentricity to spatial frequency sensitivity varies across individuals, and (b) our specific formula was based primarily on data from young adults and potentially the function for intact retinal regions in older AMD patients could be different. Also note that (c) we applied the same blur level across the whole face (i.e., we didn't blur the further-from-central-vision side of the face image more than the closer-to-central-vision side; our rationale was that AMD patients scan faces [Seiple et al., 2013], and thus would be likely to place the highest-resolution region of their intact retina over different regions of the face image in turn).

The necessarily approximate nature of our eccentricity values was one reason for testing a wide range of blurring levels. If all levels show a caricature advantage then this would provide a proof of concept that caricaturing is likely to be effective in AMD regardless of the specific eccentricity function of the individual patient, their precise damage profile (patients' scotomas are often not circular, meaning that the retina may be intact at different eccentricities in different directions; e.g., Cheung & Legge, 2005), and their precise pattern of eye movements when viewing faces.

Finally, note that our blurred stimuli can be produced from other combinations of amount-of-magnification and degree of eccentricity. For example, our 30° eccentricity stimuli for 18.11° wide are identical physically to the output of our blur simulation for 10° eccentricity with a smaller face (6.63° wide.). Thus, all references in the article to 30° or 10° should be considered shorthand for the blur produced by “30°

eccentricity at 18.11 degrees wide” and “10° eccentricity at 18.11 degrees wide.”

Procedure

Participants were tested individually. Stimuli were presented on an Apple iMac computer (Apple, Cupertino, CA) with 68.5 cm screen (resolution = 2560 × 1440 pixels; visible region of screen was 33.5 cm tall × 59.8 cm wide) running OS X using SuperLab 4.5 stimulus presentation software (Cedrus Corporation, San Pedro, CA).

Stimulus presentation

The participants viewed all images with central vision, and were allowed to make eye movements around the display and view the images for as long as they liked. The participants were positioned 75 cm from the screen using a chin rest; note that the 75 cm viewing distance produced no significant additional blurring in the stimuli (i.e., participants reported that the Blur0 faces all looked completely clear).

For presentation, items were organized into four blocks of trials, each for one of the face sets (Figure 5). In each block, each of the five individuals of that set was rated for dissimilarity to each of the other four set members. This resulted in 10 comparisons (i.e., Face A vs. Face B, A vs. C, A vs. D, A vs. E, B vs. C, etc.). Each comparison was repeated at each level of caricature (0%, 20%, 40%, 60%) and blur (0°, 10°, 20°, 30°) producing a total of 160 trials per block and 640 trials for the full experiment. The order of trials within a block was randomized for each participant, and each individual (e.g., Face A) appeared equally often on the left and the right of the screen. Order of blocks was counterbalanced: half the participants completed the two female face blocks before the two male face blocks, and vice versa for the other half of participants (the order of the face sets within each gender was the same for all participants).

To avoid similarity ratings of the two people being driven by low-level differences in face size, we also varied size of the images (Figure 4). For each person, two of the images were approximately 9 cm × 7 cm (6.9° × 5.3°), one approximately 10 cm × 8 cm (7.6° × 6.1°) and one approximately 8 cm × 6 cm (6.1° × 4.6°); note sizes are approximate because precise dimensions vary across individuals and across viewing angle. The assignment of image size and viewpoint to screen position remained consistent over trials (e.g., the largest image of the left-hand individual was always the 30° left viewpoint and appeared at the bottom-left of the screen; the largest image of the right-hand individual

was always the front-on viewpoint and appeared at the top-right of the screen).

Rating task instructions

Participants were informed that they would see multiple images of two different people, one person on the left side screen and one person on the right side. They were instructed to rate how different the two *people* looked, taking into account all four of the images presented of each person, rather than rating just how similar the *pictures* looked, on the scale of 1 (extremely similar) to 9 (extremely different). Response was via pressing the corresponding number (1–9) on the keyboard. This was followed by an intertrial interval of 300 ms.

To give room to observe differences in percepts between conditions in a rating experiment, it is important for participants to use a good range of the rating scale. To encourage this, we: (a) explicitly instructed participants to use the full scale range as far as possible rather than restricting their responses to only one part of the scale, (b) illustrated the variability in face appearance in the upcoming block of trials in a preview slide (shown for 20 s) that displayed all ten male (or female) identities presented simultaneously using a range of caricature and blur levels, and (c) told participants that some individuals might look quite a lot like others in the block but that the aim was to make graduated distinctions between *how much* different in identity the faces appeared.

Results

For each participant a dissimilarity score in each of the 16 conditions was produced, by averaging the participant's ratings across the 20 trials in that condition. Figure 6 then shows mean dissimilarity ratings across the 12 participants. As can be seen, perceived dissimilarity increased with increasing caricaturing level, and this occurred for all blur levels. Thus, there was a caricature advantage in identity perception: the two faces were perceived as more different when caricatured than when veridical. These observations were confirmed by statistical analysis as follows.

A two-way repeated measured analysis of variance (ANOVA) revealed a significant main effect of blur, Wilks' Lambda = .34, $F(3, 9) = 5.84$, $p = 0.02$ (where sphericity was violated, we report results from the multivariate approach). From Figure 6, this reflects the pattern we would expect, namely that the two faces become perceived as more similar in identity (lower dissimilarity ratings) as the blur level is increased.

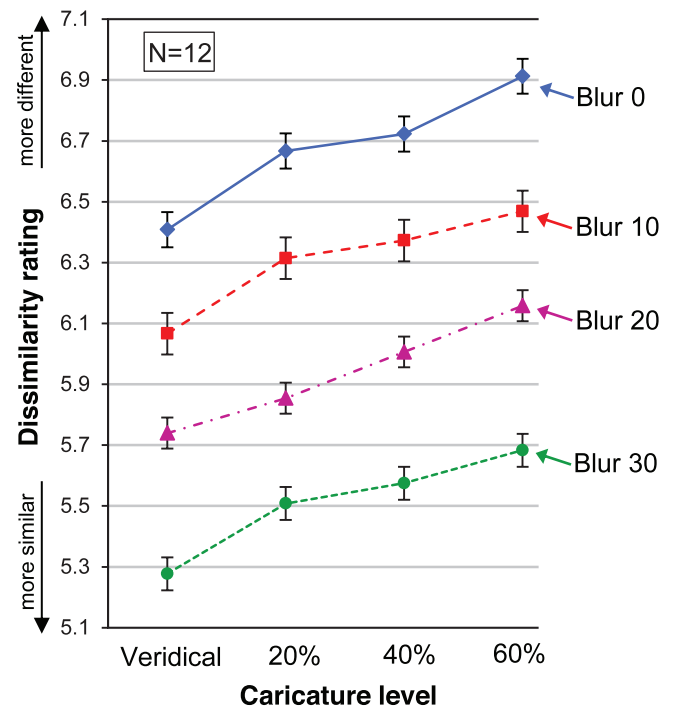


Figure 6. Experiment 1 results for the pairwise dissimilarity rating task. Results show that as faces become more caricatured, the perceived difference in identity between two faces is enhanced (i.e., rating scores increase). Also, as faces become more blurred, the two faces become perceived as less different in identity. Error bars are for the effect of caricature (i.e., ± 1 SEM derived from the MSE for the effect of caricature, at each blur level).

There was also a significant main effect of caricature, Wilks' Lambda = .09, $F(3, 9) = 28.79$, $p < 0.001$. Figure 6 shows this reflects a pattern in which the faces appear more different from each other in identity (higher dissimilarity ratings) as the images become more caricatured. There was no hint of any interaction between blur and caricature, Wilks' Lambda = .70, $F(9, 3) = .14$, $p = 0.99$, indicating that the caricature advantage had the same strength regardless of blur level.

In order to describe the exact pattern of caricature effect, we conducted trend analyses within each level of blur. For Blur0 (unblurred images reflecting normal central vision), the relationship between caricature level and increasing dissimilarity was linear, with a significant linear trend ($p < .001$) but no significant quadratic or cubic trend (all $ps > 0.13$). The same purely linear trend was found for Blur10, Blur20, and Blur30 (linear all $ps < 0.003$; quadratic and cubic all $ps > 0.3$). Thus, there was no flattening off in the rate of improvement with increasing caricature, up to the highest level tested (60% caricature level).

We were also interested in whether caricaturing was able to improve the perception of blurred faces to normal levels. Here, the question is whether any of the caricature in blur conditions produced perceived dissimilarity ratings as high as those in the veridical Blur0 condition (i.e., natural faces seen with unimpaired vision). Figure 6 shows that this was achieved for Blur10, for which the 40% caricature condition produced nearly as high a mean as veridical Blur0 (and did not differ significantly from it, $t(11) = .22$, $p = 0.83$) and the 60% condition mean slightly exceeded it (again with no significant difference, $t(11) = 0.40$, $p = 0.70$). For Blur30, caricaturing did not return identity perception to normal levels: even the 60% caricature condition remained substantially and significantly below unblurred veridical, $t(11) = 2.56$, $p = 0.03$. Together, these results show that when the face blur level simulated 10° eccentricity, caricaturing could improve the perceived differences between face identities to the level of natural unaltered faces seen with unimpaired vision. With higher levels of eccentricity, caricaturing improved identity perception but did not improve it to the level of unimpaired vision.

Finally, results were consistent across all of the four male and female face sets tested. The data were analyzed in a 4 (face set) \times 4 (caricature) \times 4 (blur) within-subjects ANOVA. There was no difference in average ratings across the four face sets, Wilks' Lambda = .61, $F(3, 9) = 1.88$, $p = 0.20$, nor did face set interact with caricature or blur (all $ps > 0.44$).

Discussion

The results of Experiment 1 demonstrate a caricature advantage for perceiving differences in face identity under low-resolution conditions. Caricatures were consistently rated as more dissimilar than veridical images, and the size of this effect was just as strong for blurred faces as for unblurred faces. With 60% caricaturing, the caricature advantage compensated for approximately 10° (technically, " 10° eccentricity at 18.11° wide") of additional blur: this restored face perception to normal levels for 10° blur. Also, it improved perception of 20° blur to the veridical 10° blur level, and perception of 30° blur to the veridical 20° level (Figure 6). Overall, results provide positive evidence arguing that caricaturing is likely to be useful in improving face identity perception in AMD.

Note that dissimilarity ratings increased linearly across the four caricature levels we tested and showed no sign of leveling off. This suggests that increasing the level of caricature further (e.g., to 80% or 100%) might perhaps continue to increase perceived dissimilarity, and further counteract the effect of blur. Unfortu-

nately, however, the practicality of making stronger caricatures is limited. Caricaturing by more than about 60% (at least with the software we used) regularly introduced problematic morphing artifacts into the image for some faces. Thus, in our face memory experiments, we use the 60% caricature level rather than trying even stronger caricatures.

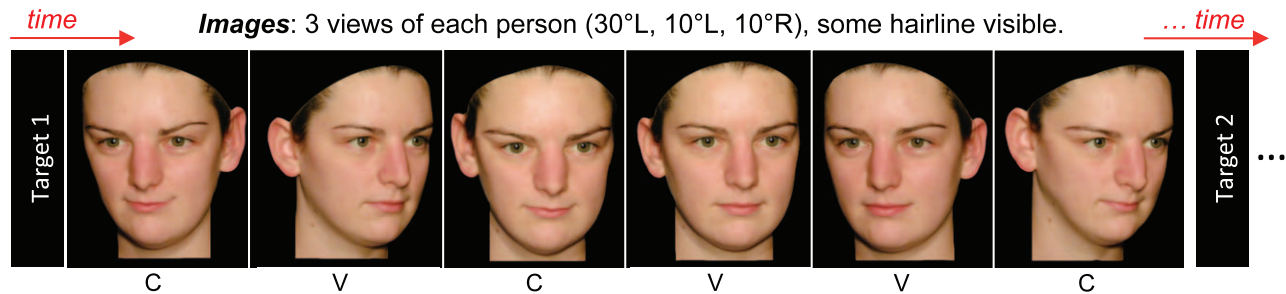
Experiments 2 and 3: Face memory

Experiments 2 and 3 tested face recognition directly in a memory task. For unblurred faces, the dissimilarity of a given face to other faces within a set (as we assessed in Experiment 1) is a strong predictor of face memory performance (Light et al., 1979).

Because face memory scores are more variable than similarity ratings, we tested more participants (e.g., following Light et al., 1979) and, further, in a 45 min testing session we were able to include fewer conditions than in our first experiment. We tested only two of the caricature levels, veridical (0%) and 60%, and we tested only three of the blur levels, unblurred (Blur0), Blur10, and Blur30. In addition, we could include only one blur level in the study phase (i.e., blur was varied only at test). We elected to use unblurred faces for the study phase. This was because we tested memory for "familiarized" faces—that is, we used an extended learning phase in which the participant sees the person in multiple viewpoints and can develop a true "face" representation of that person rather than merely learn a single image. Memory for such experimentally familiarized faces correlates well with memory for pre-experimentally familiar faces (e.g., as assessed by correlations between familiarized face performance on the Cambridge Face Memory Test and familiar face performance on Famous Face tests; Russell, Duchaine, & Nakayama, 2009; Wilmer et al., 2010). We thus see familiarized faces as a providing a reasonable approximation to the situation in which an AMD patient was familiar with a person prior to the onset of the AMD (i.e., they were learned unblurred) and then they later need to recognize that person at various stages of disease progression after onset of AMD (i.e., faces are tested at various blur levels).

In our memory task, participants first learned ten target faces, each seen in three viewpoints and repeated multiple times. In the later memory test, the ten target faces were presented among new, nontarget faces, and participants judged whether each face was one of the learned targets (old) or not (new). Because real-life face recognition requires us to recognize a person and not a specific photograph of that person or a single changeable feature (e.g., hairstyle), the images of the faces presented in the testing stage were different from

A. Learn phase of memory experiments



Procedure: • Illustration shows Learn V+C procedure (Target 1 learned in both V and C versions)

- In Learn V or C procedure, one version deleted and replaced with repeats of the other

B. Test phase images

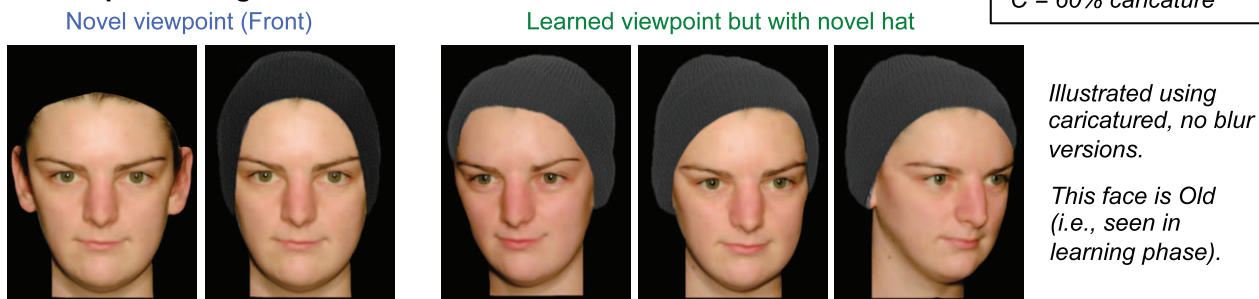


Figure 7. Memory experiments: image change between learn and test (Experiments 2 and 3), and learning procedure for the learn-both-veridical-and-caricatured [V + C] method (used in Experiment 3). (A) Learn phase images show hairline; each person was learned in three viewpoints to encourage face (not photograph) learning; and, in V + C, participants were taught that the veridical and caricatured images were of the same person (i.e., all called “Target 1”). (B) Test phase images of a studied (old) target were novel photographs of that person, i.e., either a novel viewpoint and/or with hat added. Note that apparent changes in face shape with adding the hat (e.g., front view with hat appears to have a narrower face than front view without hat) are illusory (the hat is pasted directly onto the no-hat image with no physical change to the face), and that these types of illusory changes with accessories occur in everyday life and must be generalized across by observers in order to accurately recognize people’s faces.

those in the learning stage; specifically, the test images were different in viewpoint and/or were given an added hat that covered the ears and the hairline (see Figure 7).

We examined the caricature advantage under two different learning regimes. In Experiment 2, each target face was learned *either* caricatured *or* veridical. Testing in the matched format led to two conditions for old faces, namely learn-veridical-test-veridical (VV), and learn-caricature-test-caricature (CC). For new faces, the conditions were test-phase caricature (C) and test-phase veridical (V).

In Experiment 3, each target face was learned *both* caricatured *and* veridical (see Figure 7A), meaning that V versus C status was varied only at test. Our motivation for considering this learning regime, which we code as learning [V + C], was that there is some evidence that the amount of caricaturing required to optimize face recognition can depend on the distinctiveness of the veridical face: faces that are more naturally distinctive are sometimes best recognized at a lower caricature level than those that are more

typical (Benson & Perrett, 1994; but see Rhodes et al., 1997). Thus, if caricaturing were to be implemented using a tablet computer device, learning may potentially be optimized by allowing a viewer to vary the caricature level at will to find the level they perceive as best for a given face (e.g., using a slider), effectively viewing one face at multiple levels of caricature. A potential disadvantage, however, is that learning a face in multiple formats might perhaps impede the learning process, reducing any benefit of caricaturing the face. To our knowledge, there are no previous studies testing a learn-veridical-plus-caricatured regime even for high resolution faces. To determine whether caricaturing does, or does not, assist face memory in our learn-veridical-plus-caricatured regime, we ask (a) whether a test-phase caricature advantage is present when faces are learned both as caricature and veridical, and (b) how performance compares with learning only one caricature level (from Experiment 2).

Method

Participants

All participants were Caucasian young adults recruited from the Australian National University. Experiment 2 had 31 participants (19 female, 12 male; age range 18–32 years, $M = 19.7$, $SD = 3.1$), who were recruited and tested individually and received \$15 AUD or first-year psychology course credit. Experiment 3 had 25 different participants (15 female, 10 male; age range 20–36, $M = 22.9$, $SD = 3.9$). Experiment 3 was completed as part of a third-year psychology class laboratory exercise. Participants were tested in classroom groups of approximately 10–20 people, and were given the option to have their data kept for research purposes. Each experiment took approximately 45 minutes to complete. To determine sample size, we aimed for a ballpark figure of 25 for Experiment 3 and 30 for Experiment 2 (derived from Light et al. (1979), with the larger number for Experiment 2 because the 10 “old” items in that experiment had to be split between VV and CC conditions, giving fewer items per condition than in Experiment 3). Decisions about final sample size were not influenced by looking at close-to-complete partial data.

In both experiments, visual acuity was not formally tested, but participants were included only if they reported normal vision (or corrected-to-normal wearing their usual glasses or contact lenses) on a postexperiment questionnaire, and indicated that they had no visual disorders and could see the screen clearly. Participants were also included only if they reported no history of serious head injury or disorder that might affect face recognition (e.g., epilepsy, autism spectrum disorder), and were not born more than 3 weeks premature. In Experiment 3, due to the group testing, we also gave participants the opportunity to indicate after testing whether there was any reason associated with lack of effort why we should not use their data (e.g., they were feeling ill that day).

Design

Each experiment comprised two separate stages: 10 individuals were learned and then, subsequently, a later test phase presented 20 individuals (10 old, 10 new) in an old–new decision task. In both experiments, learned faces were always unblurred, and test-phase faces (old and new) were shown at Blur0, Blur10, and Blur30. In both experiments, for old faces each individual was learned in three images (to encourage face, not photograph learning); specifically, 10° right, 10° left, and 30° left images showing ears and hairline (Figure

7A). To further ensure we were assessing genuine face memory, the test-phase images were different from the learn-phase images, showing each individual (old and new) in the 0° (front-on) view with the hairline visible, plus all four views (front, 10° right, 10° left, 30° left) with a hat added that covered the ears and the hairline (Figure 7B; note we added the hat because we did not have available different original photographs of the individuals). In both experiments, all caricatured images were at 60% caricature level, all faces were upright, and all manipulations were made within subjects (other than the manipulation of learning format between experiments).

In Experiment 2 (learn-either-caricatured-or-veridical), the 10 to-be-learned faces (Female Set 1 and Male Set 1 from Figure 5) were split into two groups of five (one with three males and two females, and the other with two males and three females). Each participant learned only veridical images for one group of five and only caricatured images for the other group, with assignment of face groups to learn-veridical versus learn-caricature condition counterbalanced across participants. At test, 10 learned faces (and the 10 new faces) were each presented in veridical format, and on separate trials in caricatured format. For old faces, the conditions of interest were: VV and CC (with the first letter coding “learn” and the second coding “test”; each with five faces per participant). New faces (Female Set 2 and Male Set 2, from Figure 5) did not appear at learning, and thus veridical versus caricature was varied only at test, giving two conditions we label as: V, C (each with 10 faces per participant).

In Experiment 3 (learn-both-caricatured-and-veridical), the only difference from Experiment 2 was that each of the 10 to-be-learned faces was seen during learning in *both* the three veridical images (10° right, 10° left, and 30° left views with hairline) *and* the three caricatured images (10° right, 10° left, and 30° left views with hairline). The participants were explicitly informed that “some images had been digitally enhanced” (no further details were given, and no mention of caricaturing was made) and that *all six images were of the same person*. Caricature condition was then varied only at test, and we label the conditions: [V + C]V and [V + C]C for old faces, and again simply V and C for (unlearned) new faces (all conditions with 10 faces per participant).

Stimuli

The 20 individuals (Figure 5) were the same as those used in Experiment 1, and the caricaturing and blurring methods were identical. The with-hat stimuli were created in Adobe Photoshop CS4 Version 11.0.2 (Adobe Systems, San Jose, CA; www.adobe.com) by

pasting a dark gray ski hat over each face, obscuring the hairline and ears, and stretching it horizontally or vertically as required to match each individual's head shape. The hat was added after caricaturing, but before the entire image was blurred.

There were 660 stimuli in total: 10 individuals \times 3 study-phase views \times 2 caricature levels (veridical, 60%) all at one blur level (Blur0) in the learning stage, plus 20 individuals \times 5 test-phase view/hat combinations \times 2 caricature levels (veridical, 60%) \times 3 blur levels (Blur0, Blur10, and Blur30) in the test stage. During the experiment, stimuli were shown at approximately 8.6° from chin to hairline and 6.2° across the widest part of the visible face, not including the ears, with approximate viewing distance 60 cm (no chinrest was used).

Procedure

Learning stage

Prior to beginning the learning stage, participants were informed that they would view ten target faces, and were asked to try to memorize these targets for a subsequent test. They were told they would view different images of each person, and that all images would show the same person even if sometimes the person looked a bit different in different photographs (as occurs in the natural world). They were also told that the photographs in the test phase would be different again, and therefore it was important to try to learn the *person*, not just a particular photograph.

In Part 1 of the learning stage, participants viewed the 10 target people in succession. For Target 1, six images were presented sequentially (Figure 7A), in random order. In Experiment 3, this comprised the face at both caricature levels (V and C) and at the three viewpoints. As participants in Experiment 2 saw the person at only one level of caricature (V or C), the three viewpoint images were presented twice. Each image appeared for 2000 ms followed by a 300 ms interstimuli interval (ISI). This procedure was then repeated for the other nine targets. The order of the targets was randomized for each participant (note that, in Experiment 2, the learn-V targets and learn-C targets were not blocked, but all 10 targets were intermixed).

In Part 2 of the learning stage, the same 60 images from the first stage were presented again, this time in completely random order. Each appeared for 2000 ms, with a 300 ms ISI in between.

The total number of trials in the learning phase was 120 (10 targets \times 6 images of each \times 2 parts). The learning phase took approximately 15 minutes.

Test stage

The test phase followed immediately after completion of the study phase. Participants were informed they would now see a series of face images, one at a time, and they were to decide if each face was a target or not as quickly and accurately as they could. They were told the faces would sometimes be blurry (as if seen from a distance) and would sometimes be “artificially enhanced,” and that it was important to judge the person, not just the photograph. Each test stimulus remained on the screen until participants made their response, by pressing the Z key on the keyboard if the face was a target or the M key if the face was new. A 300 ms ISI followed before the next trial.

The number of trials in the test phase was 600 (20 individuals \times 2 caricature levels \times 5 test-phase view/hat combinations \times 3 blur levels), presented in randomized order for each participant, and containing short breaks after trials 200 and 400. Duration of the test phase was approximately 30 minutes.

Equipment: Both experiments were run on Dell Optiplex 780 PCs (Round Rock, TX) (resolution = 1920×1080 pixels; visible region of screen $53 \text{ cm} \times 30 \text{ cm}$), using SuperLab 4.5 (Cedrus Corporation) stimulus presentation software. Screen background was black, and all face stimuli were presented at screen center.

Results

Three measures of memory performance were analyzed. For d' and accuracy (% correct for hits and correct rejections), higher scores reflect better performance, and thus a caricature advantage is reflected as caricature scores higher than veridical scores. We also examined reaction time (RT), calculated for correct responses and excluding preemptive responses (trials with RTs faster than 300 ms), and outlying values (trials with RTs greater than 2000 ms). As lower RTs indicate better performance, a caricature advantage is reflected in lower RTs for caricature than veridical.

Learn-either-caricatured-or-veridical (Experiment 2)

Figure 8 plots results for our first learning regime, in which a given face was learned either caricatured or veridical. The conditions to be compared were VV versus CC for old faces, and V versus C for new faces. Our results show that caricaturing enhanced face memory, and that, most importantly, this included under blurred conditions. Blurred caricature advan-

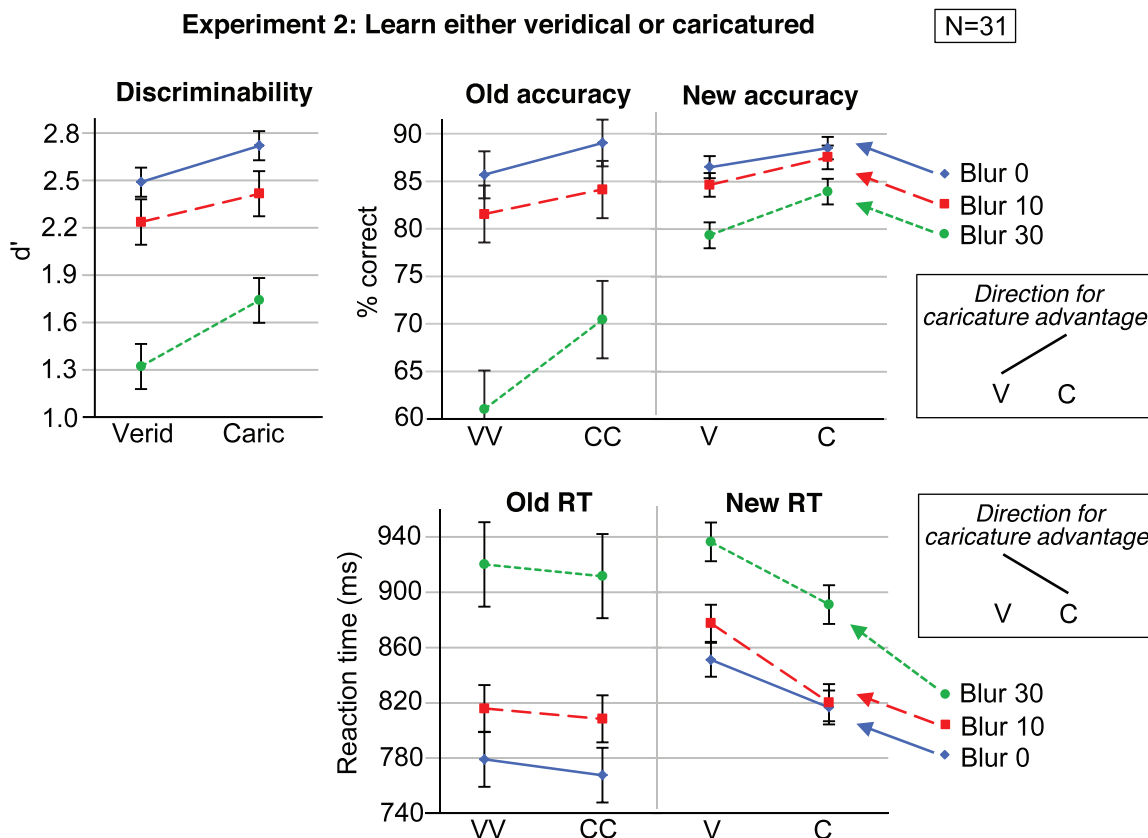


Figure 8. Memory results for Experiment 2, using the learn-each-face-either-veridical-or-caricatured regime. Scores for old faces refer to correct recognition of previously learned faces at test; scores for new faces refer to correct rejections of unlearned faces. VV = learn-and-test-the-face-veridical; CC = learn-and-test-the-face-caricatured; for new faces, V = test-phase veridical, C = test-phase caricatured. Discriminability (d') is calculated for veridical using VV (old) with V (new), and for caricatured using CC (old) with C (new). Error bars show ± 1 SEM of the difference scores for the veridical versus caricatured comparison at each blur level.

tages were also similar for old faces (Figure 8, left column) and new faces (Figure 8, right column), particularly for accuracy. These conclusions were supported by statistical analysis as follows.

First, to assess the effects of caricaturing on improving memory overall, we analyzed discriminability (d' , see Figure 8). Using a two-way within subjects ANOVA (blur level \times caricature level) we found the expected significant main effect of blur, $F(2, 60) = 95.66$, $MSE = .20$, $p < 0.001$, with discriminability decreasing with increasing blur levels. Most importantly, memory in the caricature condition was significantly better than memory in the veridical condition, with a main effect of caricaturing $F(1, 30) = 6.82$, $MSE = .52$, $p = 0.01$, and the size of the caricature advantage did not vary significantly across blur levels, demonstrated by no significant interaction between caricature and blur, $F(2, 60) = 2.00$, $MSE = .12$, $p = 0.14$. A priori comparisons conducted within each level of blur demonstrated a significant caricature advantage for Blur0 (i.e., caricatured better than veridical, $t(30) = 2.50$, $p = 0.02$) and Blur30, $t(30) = 2.93$, $p = 0.006$,

although the trend in the same direction did not reach significance for Blur10, $t(30) = 1.25$, $p = 0.22$.

We then proceeded to analyze “old” and “new” trials separately. Our motivation for this was twofold. First, our d' results have demonstrated that caricaturing improves memory for blurred faces. However, they do not tell us the extent to which this effect reflects an improvement in recognizing when people we meet are individuals we have seen before (i.e., improved hits for old faces), and/or an improvement in recognizing when people we meet are strangers (i.e., improved correct rejections of new faces). In real life, both of these abilities are equally important in order to respond to people in a socially appropriate manner: that is, when we see a person on the street, the social expectation is that we will greet them if they are familiar (old), but equally, that we will *not* greet them if they unfamiliar (new). Thus, for caricaturing to be a maximally useful method for improving real-life face recognition in AMD patients, we would ideally want to find that, relative to the veridical condition, caricaturing increases hits for old faces *and* increases correct rejections of new faces. A second motivation is that caricature

improvements can sometimes be stronger for new faces than for old faces (Kaufmann, Schulz, & Schweinberger, 2013), arguing that it is important to check the two separately.

We began by testing whether caricaturing had different benefits on old and new trials (Figure 8). We conducted blur level \times caricature level \times new/old ANOVAs for each of the two available dependent measures in turn (accuracy, RT). Results argued that caricaturing was equally beneficial for old and new faces. This was as indicated by no significant interactions involving caricature level and new/old status of the face (no two-way caricature level \times new/old interactions, accuracy: $F(1, 30) = .37$, $MSE = 239.20$, $p = 0.55$, reaction time: $F(1, 30) = 3.07$, $MSE = 10,126.32$, $p = 0.09$; no three-way caricature level \times new/old \times blur interactions, accuracy: $F(2, 60) = 1.32$, $MSE = 40.58$, $p = 0.27$, reaction time: Wilks' lambda = .95, $F(2, 29) = .78$, $p = 0.47$). That is, the amount by which caricaturing improved hits did not differ significantly from the amount by which it improved correct rejections. The ANOVAs also revealed significant main effects of caricature level (caricatured better than veridical): accuracy, $F(1, 30) = 9.84$, $MSE = 162.2$, $p = 0.004$; reaction time, $F(1, 30) = 9.47$, $MSE = 7390.17$, $p = 0.004$). There was also the expected main effect of blur level, with performance worsening as the faces became more blurred (accuracy: Wilks' lambda = .18, $F(2, 29) = 68.54$, $p < 0.001$; reaction time: Wilks' lambda = .25, $F(2, 29) = 42.66$, $p < 0.001$). Finally, the strength of the caricature advantage varied significantly across blur on accuracy (caricature level \times blur interaction, $F(2, 60) = 4.70$, $MSE = 40.47$, $p = 0.01$); although not on reaction time, $F(2, 60) = .17$, $MSE = 4119.73$, $p = 0.84$. This interaction was in a direction such that the caricaturing advantage appeared larger (not smaller) with increasing blur (Figure 8; however, note this interaction was not present in the d' analysis, and also could possibly arise merely from unblurred accuracy being closer to ceiling, and should thus not be taken as implying that the caricature advantage is necessarily stronger for blurred than unblurred faces).

Finally, we conducted a priori analyses separately for old and new faces at each blur level, to determine whether significant caricature advantages on hits and correct rejections were present in each of the simulated AMD conditions. For new faces, the caricature advantage (difference between V and C conditions in Figure 8) was significant for both levels of blurred faces on both dependent measures (Blur10 accuracy: $p = 0.03$, reaction time: $p < 0.001$; Blur30 accuracy: $p = 0.002$, reaction time: $p = 0.003$). For old faces, the caricature advantage (difference between VV and CC) was significant for the most blurred faces (Blur30), specifically on accuracy, $t(30) = 2.32$, $p = 0.03$, with the direction of the nonsignificant trend on reaction time

indicating that this did not reflect a speed-accuracy tradeoff (i.e., CC remained slightly better, not worse, than VV, $p = 0.78$). At the lower level of blur, the caricature advantage for old faces was not significant (e.g., Blur10: accuracy: $p = 0.40$, reaction time: $p = 0.66$). However, note that (a) the trends were all in the correct direction for a caricature advantage (i.e., CC trending better than VV; Figure 8), (b) the numerical size of the improvement for accuracy was at least as large for old faces as for the (significant) effect for new faces (in Figure 8, compare old with new averaged over Blur0 and Blur10), and (c) methodologically, weaker significance levels are to be expected for old faces, due to the larger error bars for old than new faces, which arise because the VV and CC conditions contained half as many faces, and thus, trials, as the V and C conditions. Regarding why we were able to find a significant caricature advantage for old at Blur30, despite the large error bars, this is likely due to overall performance being lower than in the less blurred old conditions, meaning that fewer subjects in this condition had performance approaching ceiling levels (only 10% of subjects greater than 94% in old Blur30, compared to 26% of subjects in old Blur10).

Overall, these results show that caricaturing improves memory for blurred faces. Moreover, this includes both assisting in recognition that a face has been learned before (i.e., acceptance of old faces, most clearly at 30° eccentricity), and recognition that a face has *not* been seen before (i.e., rejection of new faces, at both 10° and 30° eccentricity).

To examine how effective the caricaturing was at returning performance to normal vision, we then compared the caricature-in-blur conditions to veridical Blur0 (i.e., natural faces seen with unimpaired vision). For the 10° eccentricity level for new faces, caricaturing improved performance to slightly *above* normal-vision levels of accuracy (C Blur10 = 87.6%, V Blur0 = 86.5%; with no significant difference, $p = 0.40$) and to significantly better-than-normal RTs (C Blur10 = 820 ms, V Blur0 = 851 ms, $p = 0.005$); for old faces, accuracy was improved to nearly normal levels (C Blur10 = 84.1%, V Blur0 = 85.7%, with no significant difference, $p = 0.57$) and reaction time remained somewhat but not significantly worse than normal vision (C Blur10 = 808 ms, V Blur0 = 779 ms, $p = 0.11$). Overall, considering hits and correct rejections together, caricaturing returned performance to normal vision levels for Blur10 faces. As in Experiment 1, caricaturing did not improve the 30° eccentricity faces to normal: new faces remained significantly worse than normal vision on reaction time (C Blur30 = 891 ms, V Blur0 = 851 ms, $p = 0.01$) and approaching significantly worse on accuracy (C Blur30 = 83.9%, V Blur0 = 86.5%, $p = 0.09$); and old faces remained significant worse on both measures (accuracy: C Blur30 = 70.5%, V Blur0 =

Experiment 3: Learn both veridical and caricatured

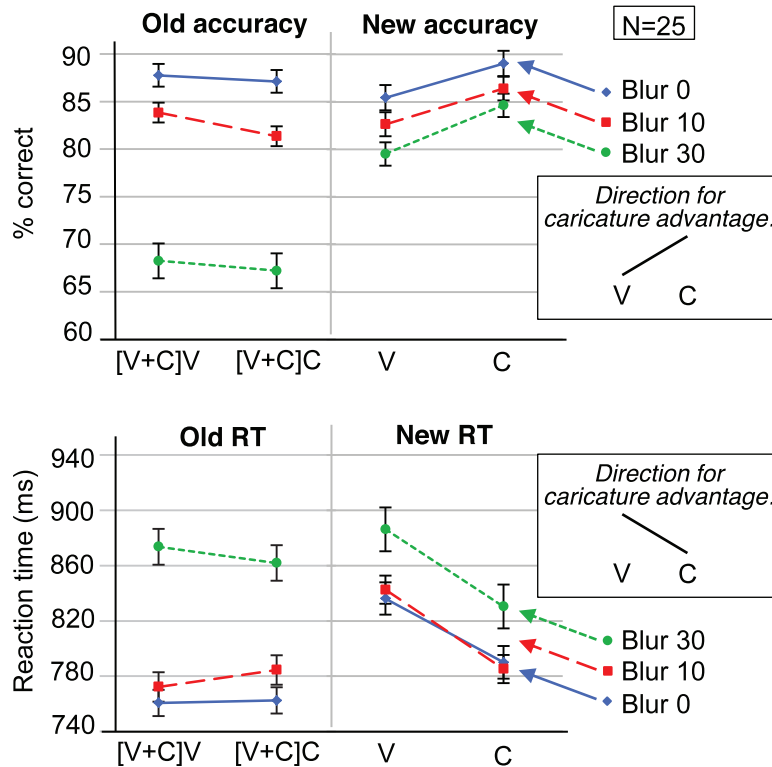


Figure 9. Memory results for Experiment 3, using the learn-each-face-both-veridical-and-caricatured regime. For old faces, [V + C]V = learn-veridical + caricatured-then-test-veridical; [V + C]C = learn-veridical + caricatured-then-test-caricatured. Error bars show ± 1 SEM of the difference scores for the veridical versus caricatured comparison at each blur level. Note the error bars are larger for old in Experiment 2 than in the other memory conditions because the VV and CC conditions contained half as many faces, and thus trials, as the other conditions.

85.7%, $p < 0.001$; reaction time: C Blur30 = 912 ms, V Blur30 = 779 ms, $p < 0.001$). Thus, as in Experiment 1, caricaturing improved blurred face recognition to normal vision levels for 10° eccentricity, but not for 30°.

Learn-both-caricatured-and-veridical (Experiment 3)

In our second learning regime—namely, one in which participants studied each learned face in both caricatured and veridical formats—results were rather different. In this case, new and old produced significantly different findings, with Figure 9 showing a caricature advantage for recognizing that new faces are not learned targets (i.e., performance in C better than in V), while there was no caricature advantage based on test-phase format for remembering old faces (i.e., performance was equal for [V + C]V and [V + C]C). That is, we found that changing the learning regime left the caricature advantage for new faces the same as in Experiment 2 (as would be expected), but removed the test-phase caricature advantage for old faces. These

observations were supported by statistical analyses as follows.

Analysis of d' (now defined using the [V + C]V and [V + C]C conditions for old trials rather than the VV and CC for old as in Experiment 2) did not show a significant caricature advantage (no main effect of caricature in a [blur \times caricature] ANOVA, $p = .17$, and also no [blur \times caricature] interaction, $p = 0.35$). However, this obscured a pattern in which caricature effects differed significantly for old and new faces, with 3-way (blur level \times caricature level \times old/new status) ANOVAs revealing significant interactions between caricature level and old–new status on both accuracy, $F(1, 24) = 16.60$, $MSE = 34.74$, $p < 0.001$, and reaction time, $F(1, 24) = 18.31$, $MSE = 2963.39$, $p < 0.001$.

Within new faces (Figure 9, right panels), we then conducted 2-way (blur \times caricature) ANOVAs for each dependent measure. As expected, increasing blur made performance significantly poorer (main effects of blur, accuracy: Wilks' Lambda = .52, $F(2, 23) = 10.72$, $p < 0.001$; reaction time: Wilks' Lambda = .70, $F(2, 23) = 4.85$, $p = 0.02$). There was also a significant advantage for caricatured faces over veridical faces (main effects

of caricature level, accuracy: $F(1, 24) = 16.95$, $MSE = 38.29$, $p < 0.001$; reaction time: $F(1, 24) = 39.84$, $MSE = 2653.29$, $p < 0.001$). The size of the caricature advantage did not vary significantly with blur (no blur level \times caricature level interaction, accuracy: $F(2, 48) = .79$, $MSE = 11.05$, $p = 0.46$, reaction time: $F(2, 48) = .26$, $MSE = 1766.12$, $p = 0.77$). A priori t tests then confirmed that the advantage of C over V was significant at each blur level considered in isolation: Blur0 ($p = 0.01$, $p < 0.001$, for accuracy and reaction time respectively), Blur10 ($p = 0.006$, $p < 0.001$), and Blur30 ($p < 0.001$, $p = 0.002$). Regarding how effective the caricaturing was at returning performance to normal levels, in the 10° eccentricity level caricaturing again improved new performance to slightly above normal levels of accuracy (C Blur10 = 86.5%, V Blur0 = 85.44%, no significant difference, $p = 0.43$) and to significantly better-than-normal reaction times (C Blur10 = 785 ms, V Blur 0 = 836 ms, $p < 0.001$). In addition, even the 30° eccentricity faces returned to very nearly normal levels (accuracy: C Blur30 = 84.6%, V Blur0 = 85.4%, $p = 0.640$; reaction time: C Blur30 = 830 ms, V Blur0 = 836 ms, $p = 0.69$), although note that this was in the context of a somewhat smaller overall blur decrement than in Experiment 2. Overall, for correct rejections, results show clear benefits of caricaturing, of similar magnitude to those in Experiment 2 (compare Figures 8 and 9).

For old faces (hits, Figure 9, left panels), in contrast, we consistently found that after learning V + C the caricature test condition, [V + C]C, did *not* lead to better memory performance than the veridical test condition, [V + C]V. Two-way (blur \times caricature) ANOVAs found the expected worsening of performance with increasing blur (significant main effects of blur, accuracy: Wilks' Lambda = .27, $F(2, 23) = 31.31$, $p < 0.001$, reaction time: Wilks' Lambda = .18, $F(2, 23) = 51.78$, $p < 0.001$). There were no interactions between caricature level and blur (accuracy: $F(2, 48) = .41$, $MSE = 28.51$, $p = 0.67$, reaction time: $F(2, 48) = 1.49$, $MSE = 1203.68$, $p = 0.24$), and there were no main effects of test-phase caricature in the direction of a caricature advantage. For accuracy, there was a numerically tiny, though statistically significant, test-phase caricature *disadvantage* (averaging over all blur levels, caricatured M in [V + C]C = 78.6% correct versus veridical M in [V + C]V = 79.9%, $F(2, 24) = 4.42$, $MSE = 16.33$, $p = 0.05$). For reaction time there was no test-phase caricature effect of any type, with means almost identical ([V + C]C = 802 ms, [V + C]V = 803 ms, $F(1, 24) = 0.01$, $MSE = 2216.68$, $p = 0.93$). A priori t tests for each blur level separately found no suggestions of any test-phase caricature advantage at any blur level or on any dependent measure (of the six t tests, smallest p for a trend in the direction of a test-phase caricature

advantage was $p = 0.37$; plus note that many conditions trended in the opposite direction, Figure 9, left panels).

Comparing learning regimes for old faces (Experiments 2 vs. 3)

So far we have reported that (a) learning each face *either* caricatured *or* veridical produces a caricature advantage for remembering old faces (CC > VV), similar in size to that for rejecting new faces, while (b) in contrast, learning each face as *both* caricatured *and* veridical leads to no test-phase caricature advantage for old faces (i.e., [V + C]C = [V + C]V). One interpretation of these findings is that, in the second learning regime, caricaturing does *not* help to improve recognition of old faces (and only improves rejection of new ones). However, this is not the only possible interpretation. Another possibility is that, by *learning* a face as caricatured and associating that caricatured form directly with the veridical form, later recognition of its *veridical* form is improved, relative to the situation in which the veridical test face has only ever been learned veridical (i.e., the VV regime).

To evaluate these interpretations we compared recognition of old faces across Experiments 2 and 3. We aimed to discriminate between two possible predictions shown in Figure 10. First, if learning a face both caricatured and veridical actually *improved* memory for the *veridical* form (i.e., caricaturing did help even in Experiment 3), then we should observe the pattern shown in Figure 10A: here, the V + C learning regime has improved memory in the veridical-test condition in comparison to the veridical-only learning regime, enhancing it to the level of both learning and testing the face caricatured (CC). Alternatively, if learning a face both caricatured and veridical removed the test-phase caricature advantage by *worsening* memory for the *caricatured* faces (i.e., caricaturing did not help in Experiment 3), then we should observe the pattern shown in Figure 10B: there, the V + C learning regime has worsened memory in the caricature-test condition in comparison to the caricature-only learning regime, taking it to the level of both learning and testing the face veridical (VV).

The actual results are plotted in Figure 10C. To maximize power in comparing across the two experiments and for efficiency of presentation, we collapsed over blur levels (see Figures 8 and 9 for means for each blur level separately) and also used inverse efficiency (Townsend & Ashby, 1983) to summarize the accuracy and reaction time aspects of performance in one measure (inverse efficiency = reaction time divided by accuracy). As can be seen, results follow the first predicted pattern, not the second. That is, [V + C]V (inverse efficiency score = 1077 ms per proportion

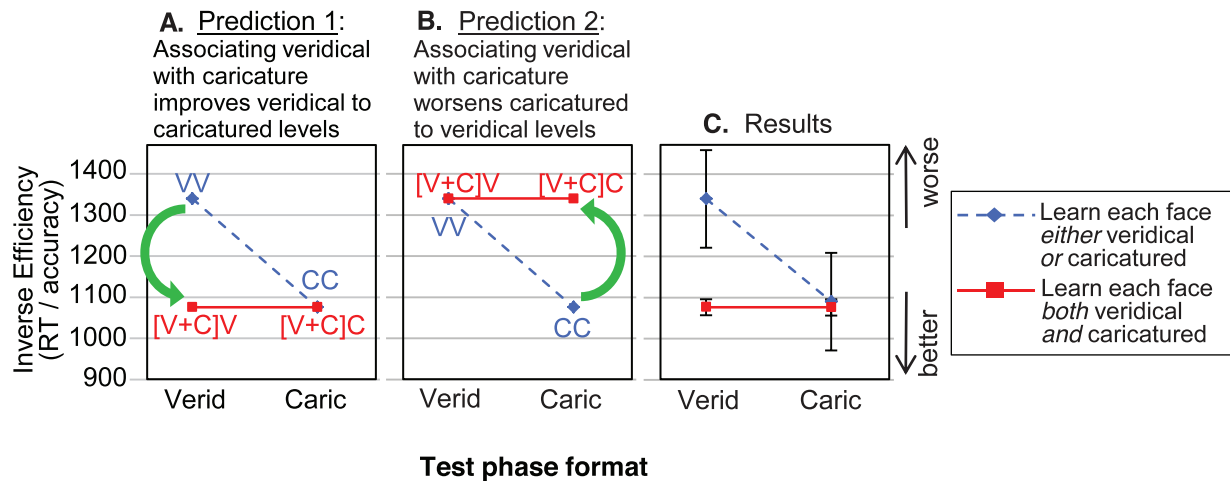


Figure 10. Comparing correct recognition of old (learned) faces across our two learning regimes: learn each face in either veridical or caricatured format (Experiment 2); or learn each face in both veridical and caricatured formats after being informed that both versions are of the same person (i.e., associating the veridical and caricatured versions at learning; Experiment 3). The dependent measure shown is *inverse efficiency*, which summarizes accuracy and reaction time (RT) together. Better performance (more accurate, shorter RT) gives a *lower* inverse efficiency score. Scores are averaged over blur level. Data for accuracy and RT separately, for each blur level, can be found in Figure 8 and 9. (A) Predicted pattern of results if the V + C learning regime improves recognition of test-phase veridical faces to that of test-phase caricatured faces. Green arrow shows predicted improvement of [V + C]V condition relative to VV condition. (B) Predicted pattern of results if the V + C learning regime worsens recognition of test-phase caricatured faces to that of test-phase veridical faces. Green arrow shows predicted worsening of [V + C]C condition relative to CC condition. (C) Results (averaged across blur), which follow the prediction in A. Error bars show ± 1 SEM of the difference scores for veridical versus caricatured test phase, suitable for the within-subjects comparison across these conditions.

correct) equaled performance in the CC condition (1090), rather than the VV condition (1340), with a significant improvement in [V + C]V compared to VV ($t(54) = 1.81, p = 0.04$, for a one-tailed direction-specific test for the predicted direction of *improvement*; note the same trend was present for accuracy and reaction time analyzed separately, although not significant in either case.)

We thus conclude that associating a veridical face with its caricatured version at learning improves later recognition of the veridical version, rather than worsens later recognition of the caricatured version. This improvement for veridical then accounts for the lack of “caricature advantage” (i.e., based on the *test* phase format) for old faces in Experiment 3, and argues that caricaturing *did*, in fact, help recognition of old faces in the learn-both-caricatured-and-veridical situation used in Experiment 3. In practical terms, this argues that a regime allowing online varying of caricature level during learning would likely be beneficial rather than detrimental for face recognition in AMD. It also suggests that allowing AMD patients to regularly study static caricatured versions of photographs of family and friends (who are also seen veridical in everyday interactions with the person) may assist the patient to recognize these individuals.

Differential sensitivity to blur for old and new faces

A finding present in both experiments, not related to caricature effects, was that blur affected memory performance more seriously for recognizing old faces than for rejecting new faces, leading to a bias to say “new” (regardless of caricature level) in the 30° blur condition. This can be seen in Figures 8 and 9, where the lines for the three different blur levels are more widely separated for old than for new, particularly affecting the most extreme blur level (Blur30) in which the accuracy for new, while still reduced by Blur30, becomes noticeably higher than for old. Statistical results from the initial global ANOVAs conducted within each experiment confirmed that both experiments showed significant blur \times old/new status interactions (Experiment 2 accuracy: Wilks’ lambda = .65, $F(2, 29) = 8.06, p = 0.002$, reaction time: Wilks’ lambda = .78, $F(2, 29) = 3.91, p = 0.03$; Experiment 3 accuracy: Wilks’ lambda = .54, $F(2, 23) = 9.76, p = 0.001$, reaction time: $F(2, 48) = 12.70, MSE = 1958.72, p < 0.001$). Thus, as faces became progressively more blurred, participants’ ability to recognize that a face had been seen before was impaired most strongly, while their ability to realize that a face was novel was impaired to a lesser extent. Because there were no three-

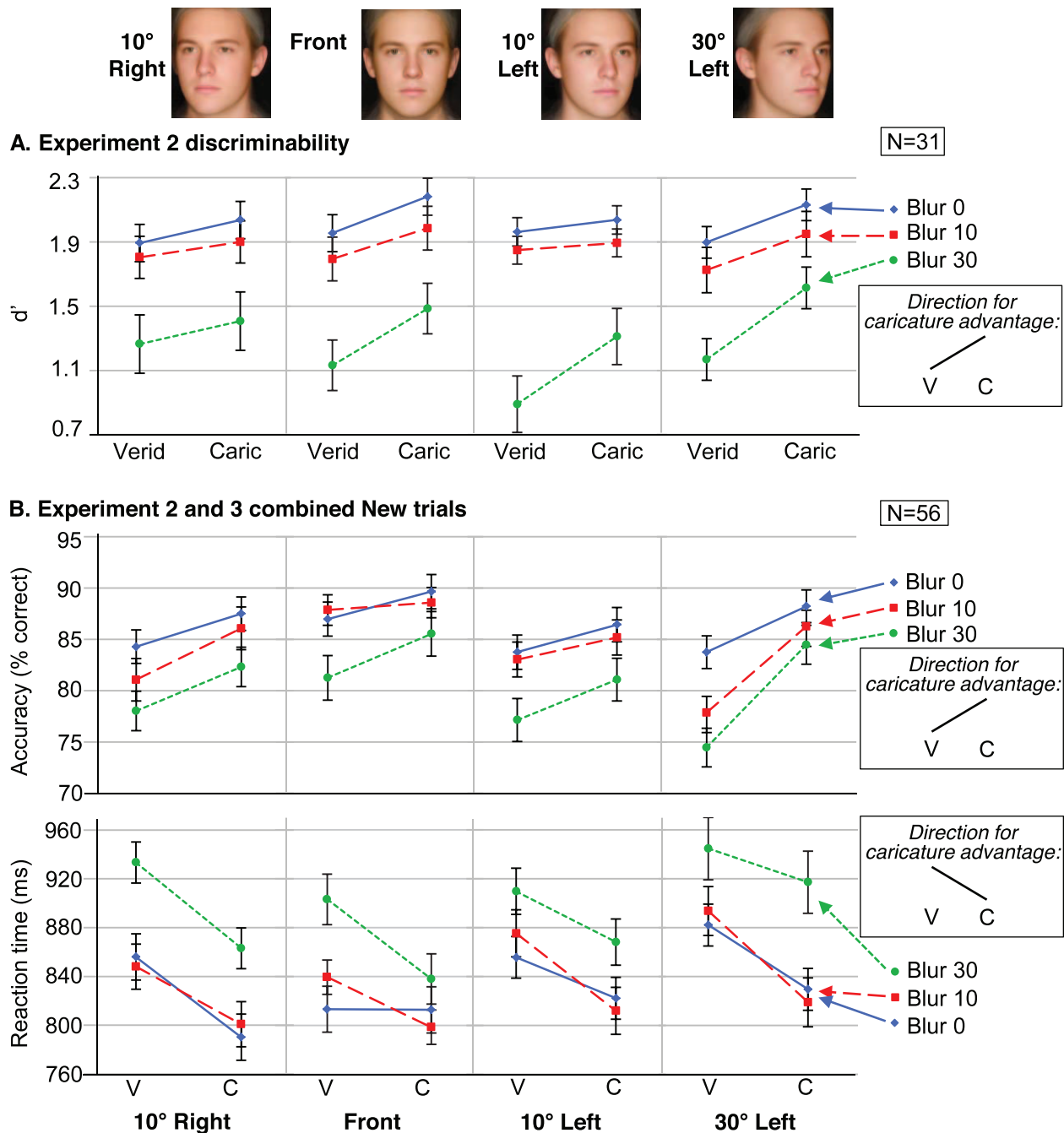


Figure 11. Caricature advantages in face memory as a function of viewpoint. (A) Results for d' from Experiment 2; note viewpoint is defined by viewpoint in the test phase (each face was learned in three viewpoints in the study phase). Veridical d' calculated using VV (old) and V (new); caricatured d' calculated using CC (old) and C (new). (B) Results for new trials only (where viewpoint can be defined independently of the differences across conditions in study-test change in viewpoint that occur for old faces, see main text), for all participants combined from Experiments 2 and 3. V = veridical; C = caricature. Error bars show ± 1 SEM of the difference scores for the veridical versus caricatured comparison at each blur level.

way interactions (blur \times old/new status \times caricature level; see earlier for statistics), this pattern was present equally for veridical and caricatured faces, and thus is not related to the caricaturing process or the caricaturing advantages.

Caricaturing and viewpoint

Returning to our core issue of caricature effects, a final question our data allowed us to address was whether caricature advantages were present at all face

viewpoints. Results (Figure 11) showed significant caricature advantages in all views, and particularly no evidence that the caricature advantage might weaken as faces were rotated away from front view. Note that for front view we analyzed the results for the with-hat stimuli, so that all four viewpoints were equivalent (i.e., all had hats).

Figure 11A shows results for d' in Experiment 2, where there was a significant caricature advantage on d' . For d' analyses, “viewpoint” was defined as the viewpoint of the face in the *test* phase (recalling that each old face was seen in several viewpoints in the study phase). Analysis showed no significant variation in the caricature advantage across viewpoint (no caricature \times viewpoint interaction, no caricature \times viewpoint \times blur interaction, all $ps > .089$). Also, significant caricature advantages were present in most viewpoints. This was revealed in caricature level \times blur ANOVAs for each viewpoint in turn, which produced significant main effects of caricatured versus veridical for three viewpoints (main effects of caricaturing, left 30° rotation $p < 0.001$; left 10° rotation, $p = 0.03$, front $p = 0.006$) and a trend in the same direction for right 10° rotation ($p = 0.16$). For d' in Experiment 3, there was also no significant variation in caricature effects across viewpoint (no caricature \times viewpoint interaction, no caricature \times viewpoint \times blur interaction; all $ps > 0.34$). Recall that the caricature advantage on d' was not significant even collapsed across viewpoint in this experiment (caricaturing at test affected “new” trials only), so we do not plot the results. The findings do, however, indicate that the lack of significant overall caricature advantage on d' in the learn-both-veridical-and-caricatured was not hiding a situation in which caricaturing improved d' for some viewpoints but not others.

A possible limitation of the above d' analysis is that it includes “old” trials and, for old trials, each face had been learned in three viewpoints at study meaning that test-phase viewpoint is intrinsically confounded with study-test *change* in viewpoint (e.g., front at test involved 10°, 10° and 30° changes from the three learned viewpoints of 10L, 10R, and 30L, while 30L at test involved 20°, 40°, and 0° changes from these three learned viewpoints). Therefore, we also conducted a viewpoint analysis using only new trials. For new trials, viewpoint can be purely defined: the face at test is simply presented in a given viewpoint. In analyzing new trials, we combined all participants from both experiments (recalling that caricature advantages for new faces were similar across experiments, see Figures 8 and 9) to maximize statistical power.

Results are shown in Figure 11B. Caricature level \times blur ANOVAs for each viewpoint in turn, all produced significant main effects of caricatured versus veridical (left 30° rotation, accuracy: $p < 0.001$, reaction time: p

< 0.001 ; left 10° rotation, accuracy: $p = 0.01$, reaction time: $p < 0.001$; front, accuracy: $p = 0.04$, reaction time: $p < 0.001$; right 10° rotation, accuracy: $p < 0.001$, reaction time: $p < 0.001$). Further, as with d' , there was no suggestion that the caricature effects were any weaker for nonfrontal views than for frontal (Figure 11B).

Discussion

Our results have confirmed that the benefits of caricaturing for AMD-type blurred faces extend to face memory tasks, assisting both in rejecting previously unseen faces as unfamiliar, and in accepting previously learned faces as known. Our results argue that these beneficial effects of caricaturing are present under learning regimes in which a given face is learned *either* caricatured *or* veridical, and under learning regimes in which a given face is learned *both* caricatured *and* veridical with the caricatured version specifically associated in memory with the veridical form (recalling that in the [V + C] regime participants were taught that both the caricatured and veridical images were *of the same person*; see Figure 7A). Our results also show that caricature advantages for blurred faces occur for all viewpoints we tested, ranging from front-view to a semiprofile (30° rotation from front). Concerning blur level, our results showed no evidence that the caricature advantages weakened at all with increasing blur level; if anything, they tended to increase in size as blur made the task more difficult and overall performance worsened. Finally, concerning returning to normal levels of face recognition (i.e., to the level of unblurred veridical faces), we found that our 60% caricature level achieved this for the 10° eccentricity condition in one experiment and very nearly even for the 30° eccentricity condition in the other (where the overall effect of blur was somewhat weaker, see Figures 8 and 9).

Overall, results of the face memory experiments, in agreement with our initial perceived dissimilarity experiment, provide a strong proof of concept that caricaturing is likely to help improve face recognition in AMD.

General discussion

Our results for AMD-type blur argue that caricaturing is likely to provide a useful method for improving face recognition under a wide variety of circumstances. We found caricature advantages across a broad range of blur levels (corresponding to 0°–30° eccentricity, blurred as for a face magnified to the size

that a patient would see on a hand-held tablet computer), for all face viewpoints from front view to semi-profile (30° rotation to the side), in perceiving differences in facial identity, and in face memory specifically including both remembering previously learned faces (old faces) and in rejecting new faces as unknown.

Importantly, we tested the individuation of faces under demanding conditions that match those required in real world recognition. That is, we (a) required observers to distinguish among several people who all look generally similar (i.e., same race, sex, and age group), (b) required observers to truly recognize the person rather than merely a particular photograph of that person by testing generalization to new images (changes in viewpoint and in whether ears and hairline were visible or covered by a hat), and (c) ensured that recognition cannot have relied on easily changeable information such as hairstyle, clothing, makeup, facial hair, or accessories such as glasses or jewelry (all of which were excluded from our stimuli). It is under these circumstances that previous studies show that high spatial frequency information, the processing of which is impaired in central vision disorders such as AMD, normally contributes importantly to face recognition (Fiorentini et al., 1983). Consistent with this importance, our observers' ability to individuate veridical faces dropped consistently with increasing blur level. However, in all cases, caricaturing improved recognition relative to the veridical and in some cases *returned face recognition to fully normal levels*—to the level of natural faces seen with unimpaired vision. This occurred for the 60% caricature level at 10° eccentricity (in pairwise similarity ratings plus both memory experiments), and even for 30° eccentricity in one case (Experiment 3) in which the overall effect of blur was somewhat weaker than in the other studies.

Advantage of a theoretical approach that includes targeting mid- and high-level vision

Our evidence that caricaturing improves recognition of blurred faces supports our theoretical position that there are potential benefits to be gained from considering the role of mid- and high-level face recognition processes in AMD, and of developing methods that enhance the stimuli in such a way that make it easier for these regions to recognize (or contribute to recognizing) the faces. We wish to emphasize that we are not arguing that these higher-level methods will necessarily be superior to previous approaches targeting low-level visual processes such as magnification alone (e.g., Tejeria et al., 2002) or increasing the contrast of the higher spatial frequencies (e.g., Peli et al., 1989; Peli, Goldstein et al., 1991). Instead, we see the low-level and

higher-level approaches as potentially complementary. That is, because each type of approach targets different stages of the visual processing stream, it may be that their benefits are additive, and thus they can be combined in the future to generate greater improvements than either method alone. This general point also applies to other higher-level aspects of face coding: for example, a method for improving holistic or part-based processing of faces could, potentially, produce additive benefits to face caricaturing.

Similarity rating provides an efficient method for measuring the caricature advantage

An important methodological finding of the present study is that the results of an identity perception task, using pairwise similarity rating (how similar in identity two faces are perceived to be; Experiment 1), match those from a direct test of face recognition (i.e., learn and remember a set of faces; Experiments 2 and 3). This is potentially useful because similarity ratings produce extremely stable data with a small number of participants, leading to both time efficiencies and a greater range of variables that can be explored in experimental research. For example, in one hour per participant, with similarity ratings we required only 12 participants to produce extremely neat data with small error bars across 16 conditions of interest (4 blur levels \times 4 caricature levels; see Figure 6). In contrast, old–new recognition with 25–31 participants produced larger error bars while testing only six conditions of interest (3 blur levels \times 2 caricature levels; e.g., Figures 8 and 9). This does not mean that testing of face memory can be avoided altogether: for example, similarity ratings alone would not have provided us with information on whether caricaturing improves both recognizing when a person has been seen before and, separately, recognizing when a person has *not* been seen before (both of which functions are important for normal social interaction). However, in exploring the space of situations under which caricaturing best enhances face individuation, researchers may find it advantageous to begin by using similarity ratings to test a large number of conditions, and then use the results to select a subset of the most interesting conditions to test fully with a face memory task.

In doing so, researchers should note an important caveat on the use of similarity ratings, namely that comparison across conditions is valid only where the manipulations have been made within participants. If different participants complete different conditions (e.g., Group 1 complete Blur0, and Group 2 complete Blur30) then their ratings scores cannot validly be compared because participants are likely to adjust their use of the scale so that their ratings cover the range of

stimuli to which they have been exposed. (This limitation also means that similarity ratings cannot be used to assess whether real AMD patients reach normal levels of face individuation performance with caricaturing, because this question requires comparison to a normal-vision control group.)

Limitations and open questions

Our present study has made a strong in-principle case that caricaturing is likely to be a useful method for assisting face recognition in AMD, but many open questions remain.

First, we tested only up to a 60% caricature level (where 100% is defined as doubling the differences of the original face from the average face). Possibly, we could improve face individuation even further by using caricature levels above the 60% maximum value we employed here. The results of Experiment 1 suggest that this might be possible: perceived dissimilarity between two faces increased *linearly* across 0, 20, 40, and 60% caricature values, suggesting that further improvements beyond 60% are likely. We note caveats, however: morphing artifacts become increasingly hard to avoid with higher caricature level (this was the primary reason for not testing above 60% in the present study); and, some studies (using only line drawn caricatures, not full photographs) have reported limits to the degree of exaggeration that can improve recognition performance even without morphing artifacts (e.g., Rhodes & Tremewan, 1994).

Second, we have tested only faces that are “familiarized.” We consider this a reasonable model of recognition of faces familiar from real life, given that the two abilities correlate fairly strongly (Russell et al., 2009; Wilmer et al., 2010); plus caricature advantages have been demonstrated for famous faces when using unblurred photographs (Benson & Perrett, 1991; Calder et al., 1996; Lee et al., 2000). However, testing caricature advantages for faces personally familiar to the participant would be valuable given the everyday importance of this task to people with AMD.

Third, we have tested memory only when faces were learned unblurred. This mimics the situation in which an AMD patient has become familiarized with a person before the onset of the disease. Concerning learning of new people *after* the onset of AMD, it would be valuable to test various degrees of blur at learning. The results from our perception experiment suggest that caricature advantages are likely to be present under these circumstances (i.e., results show that novel, unfamiliarized faces are easier to distinguish from each other when caricatured).

Fourth, the present study has tested only “own-race” faces; that is, the observers were matched in race to the face stimuli (both Caucasian). Given that in everyday life patients with AMD may see people of multiple races, it is of practical interest to know whether the caricaturing benefit is also present in “other-race” situations (i.e., where the observer is a different race from the faces they are trying to distinguish).

Fifth, we have not tested how sensitive the caricature advantage is to using average faces that are precisely matched in type to the face being caricatured. Here, we tried to use as close as possible to a perfect match: the average was matched to the to-be-caricatured face in race, sex, expression, age, and viewpoint. Our rationale for doing so was to maximize the caricaturing of *identity* information without caricaturing other aspects of the face. For example, caricaturing a Caucasian face away from an Asian average enhances the *race*-specific aspects of the face (e.g., the face is likely to become narrower, as Caucasian faces are on average more narrow than Asian faces) but may not greatly enhance the identity information in the face that distinguishes that person from other Caucasians (i.e., most Caucasian faces caricatured away from an Asian average will become narrower). The long-term aim of our project, however, is for AMD patients to be able to examine caricatures created in real time, on a tablet computer, of individuals they meet in going about their everyday lives. The method we have used here would require software that, prior to caricaturing, can automatically determine the correct average face to use: that is, to determine what race, sex, age, expression, and viewpoint category the face falls into. Currently, no method is known for fully solving this problem. Software is available to automatically locate, cut out, and expand faces from complex visual backgrounds even in video sequence (He, Kim, & Barnes, 2012), and to provide quite accurate information about the face’s sex (94% correct; Shan, 2012) and age (mean error approximately 3 years; Guo, Mu, Fu, Dyer, & Huang, 2009). However, viewpoint is less reliably estimated (70% correct within 10° error; Zhu & Ramanan, 2012), as is race (varies from 10% to 90% correct, Guo & Mu, 2010) and expression (40%–60% correct for facial action unit recognition). In total, the five-way conjunction of race \times sex \times age \times expression \times viewpoint will typically be rather poor with current methods. Thus, it would be of practical benefit if the caricature advantage were shown to survive use of the “wrong” average; for example an average matched in viewpoint to the target face, but averaged over races and sexes and with constant expression. However, note that Byatt & Rhodes (1998) found with line drawings (full photographs were not tested) that

caricaturing faces relative to wrong-race averages (e.g., Caucasian faces away from an Asian average) impaired performance and removed any caricature advantage.

Sixth, a current practical limitation on implementation of caricaturing in AMD patients, at least in a real world setting, is the need for automatic real-time caricaturing software (e.g., to allow real-time caricaturing on a tablet computer). Although we are not aware of any such software currently, some core elements needed to build it exist, specifically (a) a method for placing hundreds of landmark location points on a face and ensuring pairwise registration of these location across different frames of the face in video sequence images (in front view at least; Anderson, Stenger, & Cipolla, 2012), and (b) software for caricaturing three-dimensional heads in any viewpoint (for static images; see Jiang, Blanz, & O’Toole, 2006). However, fully automatic dense landmark localization on faces is still an active research area (e.g., it is more difficult in arbitrary pose; Anderson et al., 2012), as is its real-time computation.

Seventh, we have examined caricaturing effects only on face identification. AMD also impairs ability to perceive facial *expression*. Caricaturing of expression can increase perceived differences between expressions in unblurred photographs (Calder et al., 2000), suggesting that it may also benefit observers’ perception of expression in AMD-type blur. An important open question then concerns combining identity and expression (plus other face movements such as speech-related changes). For example, if all these aspects of a face are caricatured simultaneously, does the benefit in *identifying* the face remain?

Finally, our present research has been limited to a necessarily approximate simulation of AMD. It is not possible to precisely simulate either the degree of blur present in any individual AMD patient, nor to simulate in normal-vision observers other aspects of the way AMD patients process faces (e.g., extensive neural reorganization of peripheral visual inputs into what is normally central vision retinotopic cortex). Given this latter difficulty, we do not feel that there is benefit to be gained from trying to simulate AMD face processing slightly more precisely than we have done in the present study. Instead, we believe it is now appropriate to move to testing actual AMD patients, using our present results to guide us (e.g., by showing us that we can expect good power to see caricature advantages with only small participant numbers if we use pairwise similarity ratings). One encouraging observation from our present study is the constancy of our caricature advantage over such a wide range of blur levels. This finding argues that the wide individual patient variation in damage pattern to the retina should not significantly

affect the extent by which caricaturing can benefit face recognition in AMD.

Conclusion

The present article has made a general theoretical case that utilizing knowledge about mid- and high-level visual processing of faces can suggest new types of face image enhancements, not previously suggested by considering only early-stage visual processing, that can improve face recognition for blurred faces. Here, we have considered one such enhancement—caricaturing, derived from face-space theory—but potentially other forms of mid- to high-level enhancement could provide additional benefit (e.g., image manipulations designed to maximize strength of holistic processing of the faces while at the same time providing as much magnification of detail in local parts as possible).

While we have focused our present article on AMD, we note that potential benefits of the caricaturing method are not limited to this disorder. Other rarer diseases also cause damage to central vision and leave only peripheral vision (e.g., Stargardt’s macular dystrophy, which affects 1 in 10,000 children; Blacharski, 1988), and these diseases could also benefit from caricaturing.

Finally, our results for one form of low-resolution image—blurred with a Gaussian filter—raise the possibility that caricaturing may provide a practical method for improving face recognition in other situations in which observers see different types of low-resolution images. This includes normal vision observers watching CCTV footage (where the images are pixelated rather than blurred), and in individuals with prosthetic eye implants (i.e., bionic eyes, where images are a low-resolution display of spaced phosphenes).

Keywords: face recognition, age-related macular degeneration (AMD), caricature, face-space, blur

Acknowledgments

Thanks to Alexandra Boeing for help with using FantaMorph, and Emma Cumming for testing some participants in Experiment 2 and some preliminary data extraction in Experiments 2 and 3. Funding from Australian Research Council (ARC) Queen Elizabeth II Fellowship to EM (DP0984558); ARC Centre of Excellence in Cognition and Its Disorders (CE110001021); Australian Government as represented by Department of Broadband, Communications, and the Digital Economy; ARC Information and Commu-

nication Technologies Centre of Excellence Program; ARC Special Research Initiative in Bionic Vision Science and Technology grant to Bionic Vision Australia.

Commercial relationships: none.

Corresponding author: Elinor McKone.

Email: elinor.mckone@anu.edu.au.

Address: Research School of Psychology, Australian National University, Canberra, Australian Capital Territory, Australia.

Footnotes

¹An additional three participants were tested but their data were not used because their ratings had strong ceiling effects (they gave ratings of 8 or 9 out of 9 for every trial).

²Technical details for defining the caricature level within FantaMorph (Abrosoft Co.) were as follows. The feature curve was set to follow the features of the target face and the track curve was doubled in length. This places the average face at 0 on the scale, a 100% caricature (defined as a morph that doubles the differences between the average and the veridical face) at 100, and the veridical face at 50. To make a 20% caricature, the face extracted is that which is 20% of the distance from veridical to the 100% caricature: this falls at the 60 value on the scale, i.e., (50 for veridical) + (20% of the 50 between veridical and 100% caricature) = 50 + 10. Similarly, a 40% caricature has a scale value of 70, and a 60% caricature has a scale value of 80.

References

- Access Economics. (2010). *The global economic cost of visual impairment*. Report prepared for AMD Alliance International. Retrieved from http://www.amdalliance.org/user_files/documents/Global%20cost%20of%20VI_FINAL%20report.pdf
- Anderson, R., Stenger, B., & Cipolla, R. (2012). Dense active appearance models using a bounded diameter minimum spanning tree. In R. Bowden, R. Collomosse, & K. Mikolajczyk (Eds.), *Proceedings of the British Machine Vision Conference 2012* (pp. 131.1–131.11). Manchester, UK: BMVA Press. doi:10.5244/C.26.131.
- Anderson, R. S., & Thibos, L. N. (1999). Relationship between acuity for gratings and for tumbling-E letters in peripheral vision. *Journal of the Optical Society of America A*, *16*, 2321–2333. doi:10.1364/JOSAA.16.002321.
- Anstis, S. M. (1974). Letter: A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14*, 589–592. doi:10.1016/0042-6989(74)90049-2.
- Baker, C. I., Dilks, D. D., Peli, E., & Kanwisher, N. (2008). Reorganization of visual processing in macular degeneration: Replication and clues about the role of foveal loss. *Vision Research*, *48*, 1910–1919. doi:10.1016/j.visres.2008.05.020.
- Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Psychology*, *3*, 105–135. doi:10.1080/09541449108406222.
- Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, *23*, 75–93. doi:10.1068/p230075.
- Blacharski, P. A. (1988). Fundus flavimaculatus. In D. A. Newsome (Ed.), *Retinal dystrophies and degenerations* (pp. 135–159). New York: Raven Press.
- Blank, I., & Yovel, G. (2011). The structure of face-space is tolerant to lighting and viewpoint transformations. *Journal of Vision*, *11*(8), 15, 1–13, <http://www.journalofvision.org/content/11/8/15>, doi:10.1167/11.8.15. [PubMed] [Article]
- Brody, B. L., Gamst, A. C., Williams, R. A., Smith, A. R., Lau, P. W., Dolnak, D. . . . Brown, S. I. (2001). Depression, visual acuity, comorbidity, and disability associated with age-related macular degeneration. *Ophthalmology*, *108*, 1893–1900. doi:10.1016/S0161-6420(01)00754-0.
- Bullimore, M. A., Bailey, I. L., & Wacker, R. T. (1991). Face recognition in age-related maculopathy. *Investigative Ophthalmology & Visual Science*, *32*, 2020–2029, <http://www.iovs.org/content/32/7/2020>. [PubMed] [Article]
- Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*, 2455–2468. doi:10.1016/S0042-6989(97)00469-0.
- Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J., & Perrett, D. I. (2000). Caricaturing facial expressions. *Cognition*, *76*, 105–146. doi:10.1016/S0010-0277(00)00074-3.
- Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. I. (1996). Self-priming from distinctive and caricatured faces. *British Journal of Psychology*, *87*, 141–162. doi:10.1111/j.20448295.1196.tb02581.x.
- Chang, P. P. W., Levine, S. C., & Benson, P. J. (2002).

- Children's recognition of caricatures. *Developmental Psychology*, 38, 1038–1051. doi: 10.1.1.102.6812.
- Cheung S. H., & Legge, G. E. (2005). Functional and cortical adaptations to central vision loss. *Visual Neuroscience*, 22, 187–201. doi:10.1017/S0952523805222071.
- Crossland, M. D., Culham, L. E., Kabanarou, S. A., & Rubin, G. S. (2005). Preferred retinal locus development in patients with macular disease. *Ophthalmology*, 112, 1579–1585. doi:10.1016/j.optha.2005.03.027.
- Dagnelie, G. (2008). Psychophysical evaluation for visual prosthesis. *Annual Review of Biomedical Engineering*, 10, 339–368. doi:10.1146/annurev.bioeng.10.061807.160529.
- Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS One*, 6, e28800. doi:10.1371/journal.pone.0028800.
- deGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, 126, 87–100. doi:10.1016/j.cognition.2012.09.004.
- de Jong, P. T. V. M. (2006). Age-related macular degeneration. *The New England Journal of Medicine*, 355, 1474–1485. doi:10.1056/NEJMra062326.
- Deloitte Access Economics. (2011). *Eyes on the future: A clear outlook on age-related macular degeneration*. Report prepared for the Macular Degeneration Foundation. Retrieved from <http://www.mdffoundation.com.au/LatestNews/MDFoundationDeloitteAccessEconomicsReport2011.pdf>
- Dilks, D. D., Baker, C. I., Peli, E., & Kanwisher, N. (2009). Reorganization of visual processing in macular degeneration is not specific to the “preferred retinal locus.” *The Journal of Neuroscience*, 29, 2768–2773. doi:10.1523/JNEUROSCI.5258-08.2009.
- Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, 12, 195–201. doi:10.1068/p120195.
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12, 1187–1196. doi:10.1038/nn.2363.
- Guo, G.-D., & Mu, G. A. (2010, June). *A study of large-scale ethnicity estimation with gender and age variations*. Paper presented at the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, San Francisco, CA. Retrieved from <http://ieeexplore.ieee.org.virtual.anu.edu.au/xpls/icp.jsp?arnumber=5543608>.
- Guo, G.-D., Mu, G., Fu, Y., Dyer, C., & Huang, T. (2009, September). *A study on automatic age estimation using a large database*. Paper presented at the 12th IEEE International Conference on Computer Vision, Kyoto, Japan. Retrieved from <http://ieeexplore.ieee.org.virtual.anu.edu.au/xpls/icp.jsp?arnumber=5459438>.
- He, X., Kim, J., & Barnes, N. (2012, August). *A face-based visual fixation system for prosthetic vision*. Paper presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA. Retrieved from <http://ieeexplore.ieee.org.virtual.anu.edu.au/xpls/icp.jsp?arnumber=6346590>.
- Jiang, F., Blanz, V., & O'Toole, A. J. (2006). Probing the visual representation of face with adaptation: A view from the other side of the mean. *Psychological Science*, 17, 493–500. doi:10.1111/j.1467-9280.2006.01734.x.
- Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of multidimensional face-space. *Visual Cognition*, 4, 59–67. doi:10.1080/713756748.
- Kanwisher, N., & Dilks, D. D. (2013). The functional organization of the ventral visual pathway in humans. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences*. (pp. 733–748). Cambridge, MA: The MIT Press.
- Kaufmann, J. M., Schulz, C., & Schweinberger, S. R. (2013). High and low performers differ in the use of shape information for face recognition. *Neuropsychologia*, 51, 1310–1319. doi:10.1016/j.neuropsychologia.2013.03.015.
- Kleiner, R. C., Enger, C., Alexander, M. F., & Fine, S. L. (1988). Contrast sensitivity in age-related macular degeneration. *Archives of Ophthalmology*, 106, 55–57. doi:10.1001/archoph.1988.01060130061028.
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115–121. doi:10.1111/j.1467-9280.1990.tb00079.x.
- Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness and identification: Testing the face-space framework. *Psychological Science*, 11, 379–385. doi:10.1111/1467-9280.00274.
- Lee, K. J., & Perrett, D. I. (2000). Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. *Perception*, 29, 1291–1312. doi:10.1068/p2792.
- Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the

- monkey inferotemporal cortex. *Nature*, *442*, 572–575. doi:10.1038/nature04951.
- Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology*, *3*, 295–298. doi:10.4161/cib.3.4.12093.
- Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 212–219. doi:10.1037/0278-7393.5.3.212.
- Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G., & Wong, T. Y. (2012). Age-related macular degeneration. *The Lancet*, *379*, 1728–1738. doi:10.1016/S0140-6736(12)60282-7.
- Lowe, J. B., & Rubinstein, M. P. (2000). Distance telescopes: A survey of user success. *Optometry & Vision Sciences*, *77*, 260–269. doi:10.1097/00006324-200005000-00013.
- Mandelbaum, J., & Sloan, L. L. (1947). Peripheral visual acuity with special reference to scotopic illumination. *American Journal of Ophthalmology*, *30*, 581–588.
- Marmor, D. J., & Marmor, M. F. (2010). Simulating vision with and without macular disease. *Archives of Ophthalmology*, *128*, 117–125. doi:10.1001/archophthalmol.2009.366.
- McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research*, *49*, 268–283. doi:10.1016/j.visres.2008.10.020.
- McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M. . . . Pelleg, M. (2012). A robust method of measuring other-race and other-ethnicity effects: The Cambridge Face Memory Test format. *PLoS One*, *7*, e47956. doi:10.1371/journal.pone.0047956.
- Millodot, M. (1966). Foveal and extra-foveal acuity with and without stabilized retinal images. *The British Journal of Physiological Optics*, *23*, 75–106.
- Peli, E., Goldstein, R. B., Trempe, C. L., & Arend, L. E. (1989). Image enhancement improves face recognition. In *Noninvasive Assessment of the Visual System, Technical Digest Series*. (Vol. 7, pp. 64–67). Washington, DC: Optical Society of America.
- Peli, E., Goldstein, R. B., Young, G. M., Trempe, C. L., & Buzney, S. M. (1991). Image enhancement for the visually impaired: Simulations and experimental results. *Investigative Ophthalmology & Visual Science*, *32*, 2337–2350. Retrieved from <http://www.iovs.org/content/32/8/2337>. [PubMed] [Article]
- Peli, E., Yang, J., & Goldstein, R. E. (1991). Image variance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America A*, *8*, 1762–1774. doi:10.1364/JOSAA.8.001762.
- Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face areas in early face processing. *Current Biology*, *17*, 1568–1573. doi:10.1016/j.cub.2007.07.063.
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, *19*, 473–497. doi:10.1016/0010-0285(87)90016-8.
- Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1997). Facial distinctiveness and the power of caricatures. *Perception*, *26*, 207–223. doi:10.1068/p260207.
- Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion, and the homogeneity problem. *Visual Cognition*, *1*, 275–311. doi:10.1080/13506289408402303.
- Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*, 139–253. doi:10.1080/13506285.2013.772929.
- Rovamo, J., Virsu, V., & Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*, *271*, 54–56. doi:10.1038/271054a0.
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*, 252–257. doi:10.3758/PBR.16.2.252.
- Schmier, J. K., & Halpern, M. T. (2006). Validation of the daily living tasks dependent on vision (DLTV) questionnaire in a U.S. population with age-related macular degeneration. *Ophthalmic Epidemiology*, *13*, 137–143. doi:10.1080/09286580600573049.
- Schumacher, E. H., Jacko, J. A., Primo, S. A., Main, K. L., Maloney, K. P., Kinzel, E. N., & Ginn, J. (2008). Reorganization of visual processing is related to eccentric viewing in patients with macular degeneration. *Restorative Neurology and Neuroscience*, *26*, 391–402.
- Seiple, W., Rosen, R. B., & Garcia, P. M. T. (2013). Abnormal fixation in individuals with age-related macular degeneration when viewing an image of a face. *Optometry and Vision Science*, *90*, 45–56. doi:10.1097/OPX.0b013e3182794775.

- Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, *33*, 431–437. doi:10.1016/j.patrec.2011.05.016.
- Sjöstrand, J., & Frisén, L. (1977). Contrast sensitivity in macular disease: A preliminary report. *Acta Ophthalmology*, *55*, 507–514. doi:10.1111/j.1755-3768.1977.tb06128.x.
- Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects? *British Journal of Psychology*, *86*, 127–146. doi:10.1111/j.2044-8295.1995.tb02550.x.
- Sunness, J. S., Gonzalez-Baron, J., Applegate, C. A., Bressler, N. M., Tian, Y., Hawkins, B., . . . Bergman, A. (1999). Enlargement of atrophy and visual acuity loss in the geographic atrophy form of age-related macular degeneration. *Ophthalmology*, *106*, 1768–1779. doi:10.1016/S0161-6420(99)90340-8.
- Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50*, 300–314. doi:10.1016/j.visres.2009.11.016.
- Tejeria, L., Harper, R. A., Artes, P. H., & Dickinson, C. M. (2002). Face recognition in age related macular degeneration: Perceived disability, measured disability, and performance with a bioptic device. *British Journal of Ophthalmology*, *86*, 1019–1026. doi:10.1136/bjo.86.9.1019.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161–204. doi:10.1080/14640749108400966.
- Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*, 525–535. doi:10.1068/p150525.
- van Rheede, J. J., Kennard, C., & Hicks, S. L. (2010). Simulating prosthetic vision: Optimizing the information content of a limited visual display. *Journal of Vision*, *10*(14), 32, 1–15, <http://www.journalofvision.org/content/10/14/32>, doi:10.1167/10.14.32. [PubMed] [Article]
- Wertheim, T. (1980). Peripheral visual acuity. [I. L. Dunskey, trans.], *American Journal of Optometry & Physiological Optics*, *57*, 915–924. (Original work published 1891).
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., . . . Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*, 5238–5241. doi:10.1073/pnas.0913053107.
- Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research*, *65*, 445–451. doi:10.1016/j.jpsychores.2008.03.013.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception*, *16*, 747–759. doi:10.1068/p160747.
- Zhu, X., & Ramanan, D. (2012, June). *Face detection, pose estimation, and landmark localization in the wild*. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI. Retrieved from <http://ieeexplore.ieee.org/virtual.anu.edu.au/stamp/stamp.jsp?tp=&arnumber=6248014>.