

TeleFace: Serial Reproduction of Faces Reveals a Whiteward Bias in Race Memory

Stefan Uddenberg and Brian J. Scholl
Yale University

How is race encoded into memory when viewing faces? Here we demonstrate a novel systematic bias in which our memories of faces converge on certain prioritized regions in our underlying “face space,” as they relate to perceived race. This convergence was made especially salient using a new visual variant of the method of serial reproduction: “TeleFace.” A single face was briefly presented, with its race selected from a smooth continuum between White and Black (matched for mean luminance). The observer then reproduced that face, using a slider to morph a test face along this continuum. Their response was then used as the face initially presented to the next observer, and so on down the line in each reproduction chain. White observers’ chains consistently and steadily converged onto faces significantly whiter than they had initially encountered—whiter than both the original face in the chain and the continuum’s midpoint—regardless of where chains began. Indeed, even chains beginning near the Black end of the continuum inevitably ended up well into White space. Very different patterns resulted when the same method was applied to other arbitrary face stimuli. These results highlight a systematic bias in memory for race in White observers, perhaps contributing to the more general notion in social cognition research of a ‘White default.’

Keywords: face memory, face perception, race perception, serial reproduction

Faces, Spaces, and Races

Although the underlying biological reality of race is disputed, its psychological and perceptual reality is not (Cosmides, Tooby, & Kurzban, 2003). As such, a key task for research on social perception is to determine how race is encoded into visual memory—especially in the context of what are surely the most ubiquitous and salient social stimuli of all: faces. And as is so often the case, insights into this process may come from an exploration of how it can go awry. Visual memory for race in faces, like any form of memory, is imperfect. This can arise from noisy representations, but some of this imperfection may also arise because of *biases* in memory.

In what may be the most popular framework for discussing face perception, race is characterized as a (set of) dimension(s) within a multidimensional ‘face space.’ In this framework, a bias in visual memory may arise when some particular region(s) of the space are

prioritized—such that a face representation is in effect pulled from one region in the space toward some other region. Here we explore the possibility that visual memory for faces—using race as a case study—is in effect biased toward our representation of a ‘default face.’ We aim to document the existence of such a bias, to characterize its nature and extent, and to highlight how it conflicts in a striking way with most past research on memory for race. In doing so, we also aim to demonstrate how a variant of the method of serial reproduction—here applied to visual stimuli (in what we call the ‘TeleFace’ task)—may be especially useful for revealing the nature of such ‘default’ representations.

Representing Faces

Human faces are among the most salient and important visual stimuli we encounter in our everyday lives. Yet the differences between faces are subtle, as they all share the same parts and global configuration. Great strides have been made in recent years in understanding how we perceive and remember various aspects of such stimuli in terms of an underlying multidimensional “face space” (for recent reviews see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Valentine, Lewis, & Hills, 2016). In this framework, faces are represented not as arbitrarily varying visual stimuli, but rather in terms of a specific set of continuous dimensions that collectively comprise a psychological similarity space. As such, faces that are similar to one another on a number of dimensions are located close to one another in the ‘space,’ whereas faces with dissimilar values on many dimensions are further apart. Some such dimensions may reflect relatively simple geometric properties (e.g., nose size or forehead height; e.g., Hurlbert, 2001), while others may reflect more holistic and complex patterns (e.g.,

This article was published Online First July 16, 2018.

Stefan Uddenberg and Brian J. Scholl, Department of Psychology, Yale University.

For helpful conversation and/or comments on earlier drafts, we thank Marvin Chun, Chaz Firestone, Greg McCarthy, and the members of the Yale Perception and Cognition Laboratory. For help developing the paradigm for online testing, and for general technical wizardry, we thank Emily Ward. The data presented in this article were also presented as talks at the 15th and 17th annual meetings of the Vision Sciences Society.

Correspondence concerning this article should be addressed to Stefan Uddenberg or Brian J. Scholl, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. E-mail: stefan.uddenberg@yale.edu or brian.scholl@yale.edu

gender or competence; e.g., Little, DeBruine, & Jones, 2005; Todorov, Mandisodza, Goren, & Hall, 2005), while still others may be more ineffable (e.g., Sirovich & Meytlis, 2009).

A face is then represented in such a framework as a set of points along each of those dimensions—or, in other words, as a single point in a multidimensional space—with the distance between any two points serving to determine how discriminable those faces will be. In determining such distances, however, the dimensions themselves needn't be coded in a uniform manner. The distance between two points/faces may in part be a function of the difference in their values along that dimension (an objective property of the stimuli—e.g., how much their forehead heights differ). But the distance between two faces may also depend on the possibly nonlinear structure of the relevant continuum itself: certain regions of the dimension may be effectively stretched or contracted, perhaps a result of the statistical structure of faces one has experienced. This could then help to explain, for example, why two faces that are less familiar (e.g., of outgroup members) may be more confusable than the objectively equivalent pair of familiar faces (e.g., ingroup faces; e.g., Valentine, 1991): even though each pair may be equated in terms of some measure of objective (dis)similarity, the outgroup region of that dimension of face space may be effectively contracted relative to the ingroup region, so that the relevant points are closer together and thus more confusable.

Indeed, in most such models, those regions near the center of the face space (representing the 'norm,' or the central tendency of all previously encountered faces) are typically thought to be represented at a finer (i.e., more 'stretched') grain of resolution. This also means that different people may have differently structured face spaces, as the space effectively warps based on their own visual experience so as to best optimize their ability to discriminate the kinds of faces they actually encounter (e.g., Valentine et al., 2016). And the entire face space itself could even effectively expand or contract based on the sheer volume of faces one encounters (e.g., Balas & Saville, 2015), or could change based on the quality of contact with those faces (e.g., Walker & Hewstone, 2006; Wang & Zhou, 2016).

Several lines of empirical evidence have provided support for the existence of face spaces as the representational basis of both the perception and recognition of faces. In terms of face perception, studies of adaptation and subsequent after-effects provide evidence for coding along particular continuously varying features. If you view a masculine face for an extended period, for example, a subsequently presented novel androgynous face is biased to appear female, and vice versa (Webster, Kaping, Mizokami, & Duhamel, 2004). Such phenomena only make sense if male and female are effectively represented as the extremes of a single global continuum (for a review of such facial adaptation studies, see Rhodes & Jaquet, 2011). Similar effects occur in terms of the recognition of particular individual faces: adapting to a particular facial identity produces a high-level nonretinotopic aftereffect opposite and away from that face (through an average face, along the various dimensions diagnostic of that identity such as facial width, coloration, lip fullness, et al.), facilitating identification of the effective "anti-face" for that identity (Leopold, O'Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006). More generally, caricatures of particular individuals' faces are notoriously easier to recognize than actual exemplars of those faces themselves; this can be difficult to explain without appeal to continuously varying

dimensions of a face space (so that the relevant features that collectively define an identity can all be amplified even more than a particular exemplar; e.g., Lee, Byatt, & Rhodes, 2000). Finally, recent work in primate neurophysiology has shown that the macaque brain directly encodes face identity via a high-dimensional linear face space, whereby a given face-selective cell responds selectively to faces that vary along only a very small number of the axes of this space (Chang & Tsao, 2017).

Perceiving and Representing Race in Faces

Several previous effects involving race can be readily accounted for in terms of the notion of a face space. Perhaps the most powerful demonstration of race's impact on perception and memory, for example, comes from the other-race effect (ORE)—otherwise known as the own-race bias. The ORE is that phenomenon whereby faces of one's own race are recognized and remembered more accurately than faces of other races (for a review see Meissner & Brigham, 2001). No one race appears immune to the ORE (e.g., Bothwell, Brigham, & Malpass, 1989; Ng & Lindsay, 1994), although some evidence suggests this bias is particularly extreme in White observers (Anthony, Copper, & Mullen, 1992; Meissner & Brigham, 2001).

Many mechanisms have been adduced to account for the ORE. For example, some social-cognitive theories of race attribute the ORE to reduced motivation to individuate out-group members (e.g., Hugenberg, Young, Bernstein, & Sacco, 2010; Young, Hugenberg, Bernstein, & Sacco, 2012). This is made especially salient in cases where observers' memories for faces are differentially accurate for identical (racially ambiguous) faces, based only on how the faces are labeled (with better performance observed for own-race vs. other-race labels; e.g., Hourihan, Fraundorf, & Benjamin, 2013; MacLin & Malpass, 2001; Pauker et al., 2009). These effects can even potentially be reversed depending on subject motivation, as when angry out-group faces are more easily differentiated than angry in-group faces (Ackerman et al., 2006). And recognition for faces of one's own race can be significantly reduced by embedding the task in an interracial context, where racial identification is especially salient (Young, Hugenberg, Bernstein, & Sacco, 2009; but see Rhodes, Locke, Ewing, & Evangelista, 2009). Of course, the ORE may also be modulated by other factors, such as the strength of implicit racial biases (e.g., Walker & Hewstone, 2008) or differential levels of holistic processing between own-race and other-race faces (e.g., Tanaka, Kiefer, & Bukach, 2004).

Arguably, however, the most influential explanation of the ORE to date is informed by the notion of a face space (Meissner & Brigham, 2001). By this account, the ORE is a byproduct of the unbalanced diet of faces one consumes in daily life, as this results in a face space centered on a norm which is weighted toward, and optimized for representing faces of, the majority race (e.g., Chiroro & Valentine, 1995; Michel, Rossion, Han, Chung, & Caldara, 2006; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005).

Compelling evidence for the representation of race as a (set of) dimension(s) within multidimensional face space has come in the form of perceptual adaptation studies, which have revealed high-level aftereffects for race and other social dimensions, such as gender and emotional expression (e.g., Webster et al., 2004; for a

review see Webster & MacLeod, 2011). Additionally, other studies suggest separate neural coding mechanisms (and norms) for faces of different races, as figural adaptation aftereffects (such as those observed when faces are expanded or contracted along the vertical or horizontal axes) in faces of one race do not transfer to faces of a different race (e.g., Little, DeBruine, Jones, & Waitt, 2008; for a review see Rhodes & Jaquet, 2011).

Whereas phenomena such as the ORE pertain to the overall discriminability of faces of different races, we are interested here in the existence of systematic *biases* in memory for race—and in particular by the possibility that our memories of faces *converge* on certain prioritized regions of face space as they relate to race. This could potentially be demonstrated by exploring small biases in the memories of many different exemplar faces, as averaged across many observers. Here, though, we make this convergence especially salient by studying face memory via a novel variant of the method of serial reproduction—“TeleFace”—in which each observer is tested with only a single judgment, for a single face.

Serial Reproduction as a Method for Revealing Mental Defaults

The method of serial reproduction is a storied yet underused method akin to the children’s game of “(broken) telephone,” in which some starting message or stimulus is relayed sequentially from person to person, and morphs along the way because of the limitations of each individual’s perception and memory. In his seminal work, Bartlett (1932) used this method to explore the breakdown in memories as they are communicated, and found that given a long enough transmission chain, the final output can be very different from the initial stimulus encountered by the first observer, thanks to the accumulation of errors over time. Critically, the resulting deviations are not random, but instead are systematically related to the shared inductive biases that are collectively brought to bear by the observers: as the transmission of any given ‘message’ is corrupted (either by faulty transmission or imperfect memory), the resulting chains of reproduction will tend to converge on whatever the observers collectively think the message is *most likely* to be—with the message then remaining roughly constant after the convergence, largely regardless of how much longer the chains continue (Griffiths, Christian, & Kalish, 2006; Kalish, Griffiths, & Lewandowsky, 2007; Xu & Griffiths, 2010). In one line of studies, for example, Bartlett had subjects study and recreate an only vaguely face-like drawing accompanied by the label “Portrait d’homme.” By the end of the reproduction chain, the image had transformed into a full-fledged and easily recognizable face. In this sense, the “noise” injected by the participants in their responses is precisely the “signal” the experimenter is looking for—and so long as participants share the same inductive assumptions, their errors will be systematically biased in the direction of those assumptions. And the method of serial reproduction can then reveal these shared assumptions without ever asking any individual participant about them explicitly.

Subsequent research (especially in the literature on cultural transmission) has often made use of the method of serial reproduction (e.g., Bangerter, 2000; Lyons & Kashima, 2001; Mesoudi, Whiten, & Dunbar, 2006). And though some contemporary work has highlighted the dubious rigor some of Bartlett’s initial studies

(including the one described above; see Carbon & Albrecht, 2012), the key insight from his work remains: memory is a reconstructive process, such that mental defaults can be revealed by exploring how memories transform serially over time.

This process is especially well illustrated by one particularly inventive recent study, in which subjects were implicitly taught a function relating two arbitrary variables (Kalish et al., 2007). The initial function could represent a (positive or negative) linear relationship, a quadratic relationship, or even just random noise. After training on a given function, the first subject in a chain was then asked to reproduce it, and their reproduction was subsequently used as the function on which the next subject was trained, and so on down the line. Strikingly, reproduction chains rapidly converged to a positive linear relationship, regardless of the initial input, and even when the input was pure random noise (or even the completely opposite pattern, in the case of the negative linear relationship). In terms of arbitrary functions relating two variables, this study effectively revealed the existence of a strong positive-linear mental default.¹

Though the method of serial reproduction has seen a great resurgence lately, to our knowledge it has almost never been applied to questions about visual processing and has never been applied to face perception, per se. The present study thus attempts a novel instantiation of this method in the study of visual perception, focusing our memories for faces of different races. We hasten to add that serial reproduction may best be characterized as a ‘stylistic variant’ of more traditional memory studies, rather than having any unique ability to answer such questions. (For example, one could study face memory via reproduction without the ‘serial’ component, by testing the reproductions of many faces in many observers, and then characterizing the possible small biases that exist in many different regions of the face space.) Here, though, we think that this variant may be especially well suited to directly revealing the kinds of possible *convergence* that may exist in our memory for race—as in the case of a sort of ‘default face’ with a ‘default race.’

The Current Experiments

In the eight experiments reported below, we explore the possibility that face memory is systematically biased when it comes to race—focusing in particular on a dimension of race from (luminance-matched) Black to White. In the context of serial reproduction, such biases may be revealed by the patterns of convergence in the serial chains. If there are no such biases, then the chains may not vary systematically except as a function of their starting points: if we graph such results vertically, with time moving up (as in the results reported below), then perfect unbiased reproduction chains would simply be vertical lines (as in Figure 1a) and noisy but still unbiased chains would simply

¹ This particular experiment, like several other recent studies, was interpreted in a Bayesian framework, with the resulting mental default identified with the relevant prior probability distribution (see also Canini, Griffiths, Vanpaemel, & Kalish, 2014; Xu & Griffiths, 2010). Although our present study is fully consistent with such an interpretive framework, it does not require it. (And, by the same token, the results of Kalish et al. [2007] are just as informative and inspiring without the Bayesian packaging.) Given that none of our claims in the present article involve or turn on anything specific in the Bayesian approach—and given that the promise of this approach remains controversial (e.g. Marcus & Davis, 2013, 2015; Goodman et al., 2015), we do not use the language of Bayesian priors when discussing our results.

Possible Patterns of Convergence

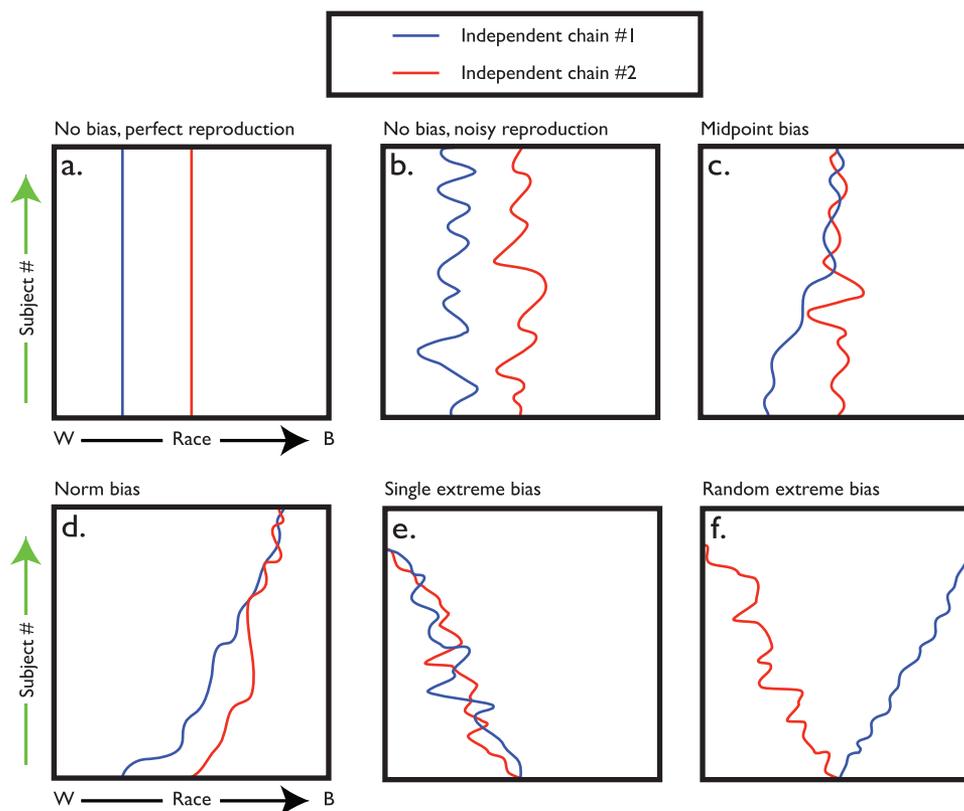


Figure 1. Several possible patterns of results for the serial reproduction of faces in a race-related continuum. For every panel, the horizontal axis represents the race of the face along a continuum from White to Black, while the vertical axis represents the participant number, such that later participants are higher up on the axis. Each panel demonstrates the potential outcomes for two independent chains of reproduction (separately drawn in red and blue, and which may vary in their starting point). See the text for details. See the online article for the color version of this figure.

involve noisy variation around those vertical lines (as in Figure 1b). (Each of the panels in Figure 1 depicts a possible pattern of results—corresponding to a different sort of possible bias—in terms of two sample reproduction chains, which may or may not start from different possible starting faces.) If there is a bias in this dimension of face space toward the midpoint of the continuum, however, then the chains should drift toward the vertical center of the graphs over time (as in Figure 1c). If the bias is to a statistical norm, in contrast, then the chains should converge over time toward some different particular point along the continuum (as in Figure 1d). Finally, it is possible that any deviations from the actual or perceived midpoint may become accentuated over time, such that chains converge toward a single extreme point in the space (as in Figure 1e) or perhaps toward either possible extreme, randomly determined for each chain based on noise in the initial deviation (as in Figure 1f).

To our knowledge, no existing data entail or rule out any of these possibilities, many of which seem plausible. For example, given the salience of race-based encoding and the possible relative rarity of racially ambiguous faces in many people's

experience, it seems reasonable to expect that if a chain begins with a racially ambiguous face (i.e., near the midpoint of the continuum), then any small deviations from that midpoint may snowball over time, as each subsequent observer compounds the previous observer's error, making the resulting face less and less ambiguous over time (as in Figure 1f, e.g.). But it seems equally plausible that reproduction chains could converge over time toward some particular region of the continuum (as in Figure 1d) that serves as a sort of 'attractor.' We thus set out to answer this question, but without any strong hypotheses (or, rather, with too many strong but mutually inconsistent hypotheses!) about what the results would be.

Experiment 1: Starting in the Middle

We first explored the outcome of serial reproduction in chains of White participants, where the first participant in each chain viewed the racially ambiguous face that sat at the midpoint of the spectrum from (luminance-matched) White to

Black.² Across all of our (race-based) experiments, we utilized two different face continua, designed to look as different (and to be constructed as differently) as possible—one using computer-generated faces, and one using averaged (real) face photographs (see Figure 2). This helps to ensure that any systematic results cannot be an artifact of any particular stimulus.

Method

Participants. For each experiment, 300 naïve White U.S.-based participants were recruited using the Amazon Mechanical Turk online labor market (MTurk). (For discussion of this pool's nature and reliability, see Crump, McDonnell, & Gureckis, 2013; Germine et al., 2012). Before beginning the experiment, participants reported their MTurk worker ID and anonymously shared demographic information including race, age, gender, and nationality. Each participant completed a single trial (in a session lasting approximately 2.5 min) for modest monetary compensation, with worker ID screening ensuring that individuals could not participate more than once for any experiment using the same face continuum.

We decided before data collection began to test 30 reproduction chains of 10 participants each, for a total of 300 participants. These values were chosen arbitrarily to be roughly in line with past serial reproduction studies, and the same sample size was used for each of the 5 reproduction studies reported here. Participants were excluded for failing an online attention check (as described below), taking more than 60 s to complete the task, or being part of incomplete chains when the final sample size was reached. The participants excluded for these reasons in all studies reported in this article are detailed in Table 1.

Apparatus. The experiment was conducted using custom software written using a combination of PHP, Javascript, CSS, and HTML. Participants completed the experiment via a custom web page which could be loaded in any modern web browser on their own laptop or desktop computers. (Because our experi-

ment required a relatively large display and the use of a computer mouse to make responses, mobile devices such as phones and tablet computers were explicitly disallowed, and attempts to access the experiment from such a device led to its immediate termination along with an error message.) Before the experiment began, each participant also completed a simple browser compatibility check to ensure that their system was able to process the required sorts of animations and response measures, and only participants who passed that check were allowed to continue.

Stimuli. Depictions of the two 61-face-long continua used in this experiment can be found in Figure 2, with FaceGen faces above and averaged photographs below. For the computer-generated face continuum, White and Black faces (used as the end points of the face continuum) were first created using FaceGen Modeler 3.5 (Singular Inversions, <http://www.facegen.com/>), which allows the creation of 3D faces based on a statistical model derived from laser scans of 271 real human faces (for details see Blanz & Vetter, 1999).³ We then used FantaMorph 5.4.2 Deluxe photo morphing software (Abrosoft, <http://www.abrosoft.com>) to create a morph series between the original pair of faces, using the 112 points of correspondence provided by the Face Locator option, manually adjusted to fit the initial two faces. We created a morph series of 61 faces (including the two initial end points), such that each successive face differed by 1.67% of the difference between the end points. Each face was then converted to grayscale, and the FaceGen faces only were placed into an oval mask which rendered the hairline and ears invisible. Finally, these images were matched for mean luminance using the SHINE toolbox in MATLAB (MathWorks, www.mathworks.com), to ensure that our results could not be explained because of differences in mean luminance across the photographs (Willenbockel et al., 2010).

For the averaged photographs, the Black and White end points of the luminance-matched 61-face-long continuum were generated by averaging together 16 Black and 16 White faces respectively, arbitrarily chosen from the Chicago Face Database (Ma, Correll, &

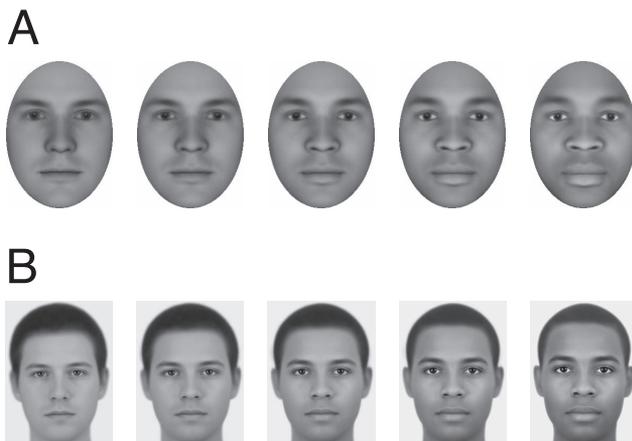


Figure 2. A depiction of the face stimuli used throughout the experiments: (A) Computer-generated FaceGen faces, and (B) Averaged photographs. Both 61-face-long continua were generated by smoothly morphing between a White face and a Black face and then matching them for mean luminance using the SHINE toolbox in MATLAB. See main text for more details.

² We matched faces for mean luminance to ensure that any systematic results are due to differences in race and not just lightness per se. Unconfounding race and lightness is made possible by the fact that morphological cues overpower lightness cues in the perception of race (e.g. Brooks & Gwinn, 2010; Gwinn & Brooks, 2015; Strom, Zebrowitz, Zhang, Bronstad, & Lee, 2012; Willenbockel, Fiset, & Tanaka, 2011).

³ In this experiment, both end point images were created under the corresponding ethnicity's tab (*European* for the White face, *African* for the Black face) nested under the *Generate* tab. *Gender* sliders were set to *Male*, *Age* sliders were set to *30*, and the *Asymmetry* slider was set to *Symmetric*. For the White end point face, the *Caricature* shape morph slider was set to *The average*, and the *Texture* morph slider was set to *Attractive*, whereas for the Black end point face, the *Shape* morph slider was set to *Attractive*, and the *Texture* morph slider was set to two notches above *Attractive*. The *Texture Gamma Correction* slider was set to 2.0 (the default value). All other sliders were allowed to reposition themselves automatically to meet these stipulations. These procedures were performed to produce faces that looked racially unambiguous without being the FaceGen average face for each race, which participants may have seen before in other MTurk experiments utilizing FaceGen faces, depending on their worker history.

Table 1
Excluded Participant Reasons (and Counts) for Each Experiment

Reason	Experiment							
	1	2	3	4	5a	5b	6a	6b
Failed attention check	44/61	36/68	53/61	28/30	11/4	1/1	57	43
RT more than 60 s	10/8	11/10	8/10	8/0	0/0	0/0	2	2
Extra participants	13/20	22/25	31/12	0/5	0/2	0/3	54	58

Note. For Experiments 1–5, the first number corresponds to the exclusion counts for participants tested with FaceGen faces, whereas the second number (after the slash) represents exclusion counts for participants tested with averaged photographs.

Wittenbrink, 2015), using PsychoMorph (Tiddeman, Burt, & Perrett, 2001).⁴ Luminance-matching and continuum-generation procedures were the same as those reported for the FaceGen faces. All faces were presented as grayscale images, presented near the center of the display (FaceGen: 202 by 284 pixels; Averaged photographs: 210 by 284 pixels) as in Figure 2.

Procedure. Participants first read detailed written instructions for the task, as described below. They then positioned their browser window such that a 1-pixel-thick circular ring (580-px diameter, drawn in black on a white background) was fully visible. After clicking on a small box labeled “Start” (which was only available after all images to be shown had been preloaded), the experiment began with a simple check to ensure that participants were attending focally to the correct region of the display: the words “PAY ATTENTION” (drawn in black, in 50-px Trebuchet MS) appeared for 4 s in the center of the ring, followed by a countdown in which the numerals 5 to 0 appeared in succession in the center of the display (in the same font), each for 1 s. Participants had been instructed that one of the numerals would be presented in red, and that they would have to report that numeral’s identity at the end of the session. (In fact the red numeral was always either “1” or “2.”)

Immediately after the offset of the “0,” the face memory task began. The face to be remembered (as described below) was presented in the center of the ring for 1 s, after which nothing but the response ring was visible for another 1 s. Then, the reproduction phase began with the presentation of an initial *probe face* (selected via a different random choice for each participant from the full morph series), shown in the same location where the initial face had appeared. A green rectangular marker (10 px by 20 px, depicted in Figure 3) also appeared at a location along the edge of the ring (selected via a different random choice for each participant). Participants then moved their mouse cursor around the ring. As they did so, the marker rotated continuously around the ring so that it was always in the location closest to the cursor, and the face presented in the center of the ring smoothly transitioned from White to Black, with that linear continuum mapped onto exactly half of the ring’s circumference (starting at a point along the ring that was selected via a different random choice for each participant). The other half of the ring’s circumference then corresponded to the exact inverse of the continuum, such that if a participant rotated the marker by 360°, the face would effectively morph smoothly throughout the entire continuum and back. Participants were instructed (ahead of time) to adjust the face (by moving the marker) until it matched the initially presented face as closely as possible—at which point they clicked their mouse to record their

selection (which turned the marker red, and made it stop tracking the mouse’s movements). (If a participant was dissatisfied with their choice at this point, they could simply click on the marker once more to turn it green and allow them to make a new selection. And no selection could be recorded until participants had moved the marker at least 50 different individual face morphs away from its random starting position.⁵) Once satisfied with their selection, participants clicked a box (presented under the ring) labeled “Next” to continue. Clicking the “Next” button marked the end of the single trial that each participant contributed to the experiment. This procedure is depicted in Figure 3.

Participants were then asked (via written prompts with attached response boxes) about (a) which countdown number had been red; (b) whether they encountered any technical problems during the experiment; and (c) “What do you think we were testing in this survey?” They recorded their responses by clicking on a box labeled “Submit,” after which they were given a confirmation code for their MTurk compensation (along with a brief description of the purpose of the experiment), and the experiment ended.

Every 10-participant reproduction chain began by showing the first participant in the chain the face from the center of the morph series (namely, face #31). For the remainder of that chain, the

⁴ To make the averaging as accurate as possible, we identified 179 landmarks on every face, and then used photo manipulation software to blur the area around the hair, ears, and neck—and to remove any remaining ‘ghostly’ traces of the original exemplar faces in those areas. The 16 neutral faces used to create the White average face were as follows (each of which had the prefix “CFD-WM-”): 003-002-N, 004-010-N, 006-002-N, 009-002-N, 010-001-N, 011-002-N, 012-001-N, 013-001-N, 014-002-N, 016-001-N, 017-002-N, 019-003-N, 024-015-N, 029-023-N, 031-003-N, and 035-032-N. The 16 neutral faces used to create the Black average face were as follows (each of which had the prefix “CFD-BM-”): 001-014-N, 002-013-N, 003-003-N, 005-003-N, 009-002-N, 010-003-N, 011-016-N, 015-015-N, 016-036-N, 022-022-N, 026-002-N, 028-002-N, 029-024-N, 032-024-N, 034-031-N, and 037-033-N.

⁵ Participants had to move through nearly the full face space before responding simply so that they could experience the full range of possible faces before making an otherwise-possibly-premature response. Because the marker’s starting point was entirely random, however, note that this could not introduce any systematic response bias. Of course, on trials where the initial marker position just happened to be close to the correct answer, this could cause the participants to initially shift their responses away from that point. However, this was always balanced out by trials in which the initial marker position happened to be far away from the correct answer, such that the required initial exploration phase caused them to move *closer* to the correct answer. (And of course, because this procedure was the same for every single condition, it cannot be responsible for any of our key results, which always involve differences across conditions.)

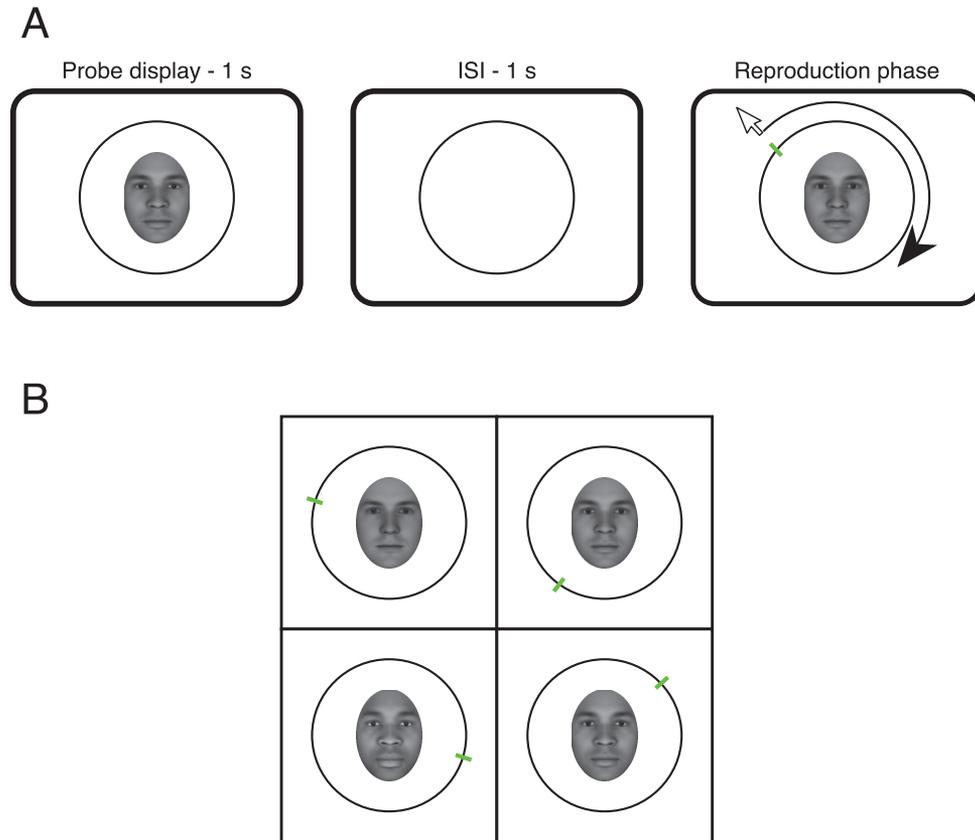


Figure 3. (A) A depiction of the reproduction task, where a participant had to reproduce an initially presented probe face by moving a marker around a circular slider to morph a response face through the face space. (B) Examples of how faces at different points in the Black/White continuum were drawn corresponding to different positions of the marker during the response phase. See the online article for the color version of this figure.

initial face shown to each participant was always the final face selected by the previous participant (though this manipulation was never made explicit to participants).⁶

Results and Discussion

Temporal analyses. The mean reproduction chains are depicted via the blue lines in Figures 4 (FaceGen faces) and 5 (averaged photographs)—where each point is simply the average of the faces selected by the 30 participants who completed that ‘step’ in their particular chains (and the where the error shading reflects the 95% confidence interval at each step in the chain). Inspection of these particular mean reproduction chains suggests two prominent patterns: (a) chains were biased White-ward, compared with the starting midpoint; and (b) chains veered into White space rapidly (but incrementally) via a few small jumps and then stayed there for the rest of the chain (continuing in the graph as a largely vertical—if noisy—line). These impressions were verified via the statistical analyses reported below and in the top section of Table 2 (for FaceGen faces) and Table 3 (for averaged photographs).

For the FaceGen faces (see Figure 4), 10 of the 10 points were White-ward of the starting midpoint face (represented here by the black vertical line), which of course is itself a significant

result by an exact binomial sign test ($p = .002$). Furthermore, a series of t tests revealed that five of the steps (#4, 7–10) differed significantly from the starting point, even after correcting for multiple comparisons ($\alpha_{\text{corrected}} = .005$, all $t_{s[29]} \geq 3.10$, $p_s \leq .005$, $d_s \geq .56$). These results indicate that there is a systematic White-ward bias—such that the final point in the mean chain was itself more than 10 faces into White space relative to the midpoint (roughly 36% of the way into White space; $M = 20.27$, $SD = 15.31$), $t(29) = 3.84$, $p < .001$, $d = .70$). However, only five of the 10 steps moved White-ward relative to the previous step, as depicted visually in Figure 4 by the number of times the blue line slopes to the left. This

⁶ Each participant was assigned the first available position in the first available chain of reproduction on a first-come, first-serve basis. When one chain was completed with 10 participants, the next participant was assigned to the first position in a new chain. However, the serial nature of this task, combined with the continuous availability of the study for MTurk workers, sometimes required running multiple chains concurrently. For example, if one participant began while another participant was already engaged in the task in a given chain, they were shifted to the next available position in a different chain, or were made to occupy the first position in a new chain constructed for them. As a result, some incomplete (or extra) chains were discarded at the moment that we reached our target of 30 complete chains, as detailed in Table 1.

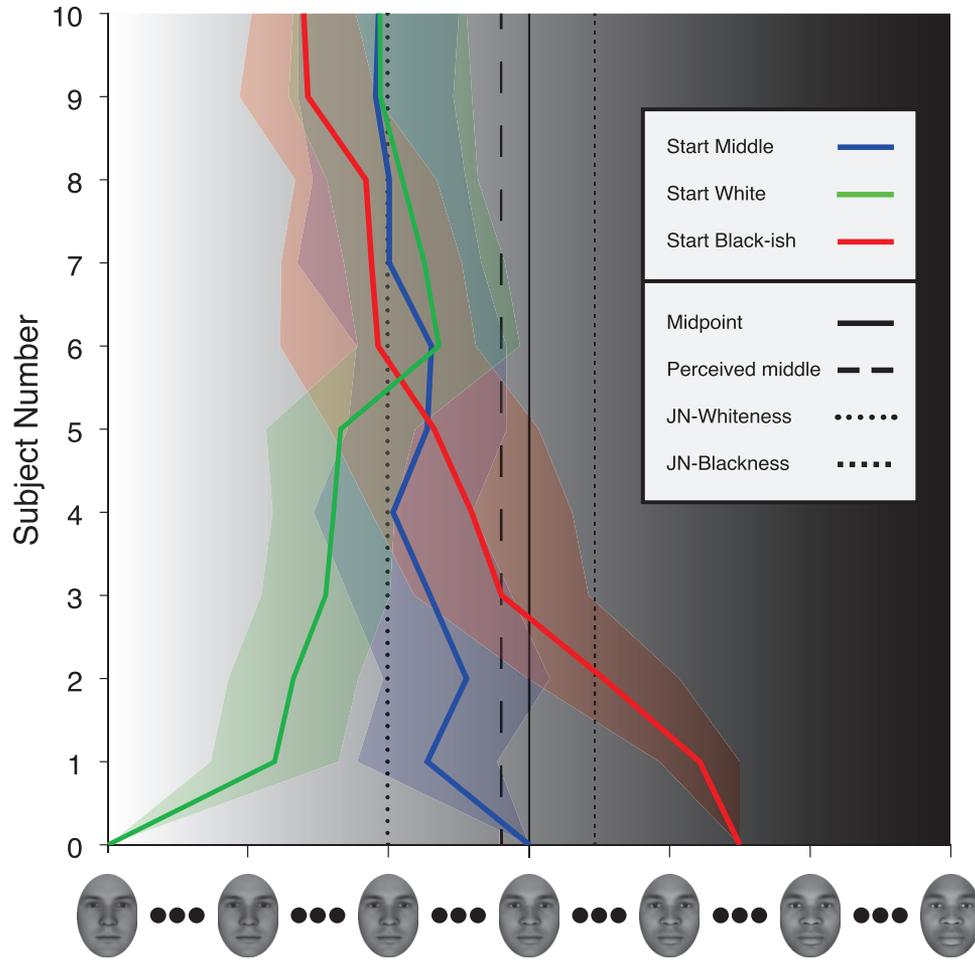


Figure 4. The mean chains of reproduction in a luminance-matched race continuum. The horizontal axis represents the face space, from White to Black. The vertical axis represents the participant number. The colored lines each represent the mean response of 30 chains of 10 participants each - when starting at the midpoint (in blue; Experiment 1), starting Black-ish (in red; Experiment 2), or starting maximally White (in green; Experiment 3). Shading represents 95% confidence intervals. The black vertical line is placed at the continuum's midpoint. The dashed vertical line is placed at the average perceived midpoint (as measured in Experiment 4). The dotted vertical lines represent the “just-noticeable Whiteness” and “just-noticeable Blackness,” as measured respectively in Experiments 5a and 5b. See the online article for the color version of this figure.

provides further evidence that there is an initial series of jumps into White space, after which the chain simply vacillates within its preferred region of the space. Moreover, a series of Bonferroni-corrected paired-samples t tests revealed that no step differed significantly from its preceding step, ($\alpha_{\text{corrected}} = .005$, all $ts[29] \leq 2.99$, $ps \geq .005$, $ds \leq .55$)—providing further evidence that individual jumps within the face space are relatively small, and that the bias Whiteward accrues with successive reproductions.

For the averaged photographs (see Figure 5), nine of the 10 points were Whiteward of the starting midpoint face ($p = .021$), and the t tests revealed that one of these steps (#6) differed significantly from the starting point even after correcting for multiple comparisons ($\alpha_{\text{corrected}} = .005$, $t[29] = 3.10$, $p = .004$, $d = .57$)—though it should be noted that seven of the steps (#4–9) differed significantly from the starting point at $p < .05$ (all $ts[29] \geq 2.06$, all $ps \leq .049$, all $ds \geq .37$).

Thus there was again a Whiteward bias, although the final point in the mean chain was itself only marginally significantly Whiter than the midpoint ($M = 25.97$, $SD = 14.02$), $t(29) = 1.97$, $p = .059$, $d = .36$. And as with the FaceGen faces, only five of the 10 steps moved Whiteward relative to the previous step. And again, a series of Bonferroni-corrected paired-samples t tests revealed that no step differed significantly from its preceding step ($\alpha_{\text{corrected}} = .005$, all $ts[29] \leq 1.37$, $ps \geq .183$, $ds \leq .25$).

Atemporal analyses. All of the conclusions we draw from our data in this article rely only on temporal analyses of the sort reported above (which we also report for each subsequent serial reproduction experiment). Because each participant only performed a single trial and was blind to the serial nature of the data collection, however, these patterns (in particular the Whiteward bias) can also be revealed in related atemporal analyses. And so although we do not report such atemporal analyses for any of the subsequent experiments, here we do

Table 2

Results of Temporal Analyses for Experiments 1–3, When Participants Viewed Luminance-Matched FaceGen Faces

Ss#	<i>M</i> (<i>SD</i>)	Versus middle		Versus perc. middle		Versus JNW	
		<i>p</i>	(<i>t</i> , <i>d</i>)	<i>p</i>	(<i>t</i> , Δ)	<i>p</i>	(<i>t</i> , Δ)
“Start Middle” chains (Experiment 1) – FaceGen faces							
1	23.73 (13.35)	.006	(2.982, .544)	.043	(2.109, .962)	.300	(1.049, .247)
2	26.53 (15.90)	.135	(1.539, .281)	.410	(.835, .451)	.079	(1.803, .492)
3	23.93 (15.58)	.019	(2.485, .454)	.090	(1.750, .926)	.330	(.987, .265)
4	21.30 (15.16)	.002	(3.505, .640)	.010	(2.730, 1.407)	.897	(.130, .034)
5	23.73 (15.20)	.014	(2.619, .478)	.072	(1.862, .962)	.353	(.941, .247)
6	24.03 (14.29)	.012	(2.670, .487)	.072	(1.863, .907)	.279	(1.096, .273)
7	21.03 (17.57)	.004	(3.108, .567)	.020	(2.449, 1.455)	.971	(.036, .011)
8	21.03 (14.61)	.001	(3.736, .682)	.006	(2.925, 1.455)	.966	(.042, .011)
9	20.07 (14.81)	<.001	(4.043, .738)	.003	(3.238, 1.632)	.775	(.287, .074)
10	20.27 (15.31)	.001	(3.840, .701)	.004	(3.066, 1.596)	.832	(.213, .056)
“Start Black-ish” chains (Experiment 2) – FaceGen faces							
1	43.13 (7.68)	<.001	(8.656, 1.580)	<.001	(9.392, 2.582)	<.001	(12.291, 1.946)
2	36.30 (14.45)	.054	(2.009, .367)	.011	(2.710, 1.334)	<.001	(5.354, 1.347)
3	29.03 (16.63)	.522	(.648, .118)	.991	(.011, .006)	.017	(2.505, .711)
4	26.90 (19.16)	.251	(1.172, .214)	.558	(.593, .384)	.113	(1.627, .524)
5	24.23 (19.93)	.073	(1.859, .339)	.205	(1.295, .871)	.390	(.871, .291)
6	20.23 (18.67)	.004	(3.158, .577)	.016	(2.539, 1.602)	.852	(.188, .059)
7	19.77 (17.22)	.001	(3.573, .652)	.007	(2.893, 1.687)	.734	(.342, .100)
8	19.37 (13.50)	<.001	(4.718, .861)	.001	(3.814, 1.760)	.573	(.568, .135)
9	15.23 (13.06)	<.001	(6.614, 1.208)	<.001	(5.629, 2.515)	.037	(2.148, .497)
10	14.93 (9.92)	<.001	(8.868, 1.619)	<.001	(7.432, 2.570)	.007	(2.791, .523)
“Start White” chains (Experiment 3) – FaceGen faces							
1	12.87 (12.19)	<.001	(8.149, 1.488)	<.001	(7.041, 2.948)	.002	(3.216, .704)
2	14.20 (12.39)	<.001	(7.429, 1.356)	<.001	(6.361, 2.704)	.011	(2.649, .587)
3	16.50 (12.27)	<.001	(6.472, 1.182)	<.001	(5.420, 2.284)	.086	(1.754, .386)
4	17.07 (11.65)	<.001	(6.551, 1.196)	<.001	(5.433, 2.180)	.118	(1.592, .337)
5	17.57 (14.20)	<.001	(5.182, .946)	<.001	(4.316, 2.089)	.245	(1.180, .293)
6	24.57 (15.48)	.030	(2.276, .415)	.134	(1.540, .810)	.238	(1.199, .320)
7	23.53 (15.37)	.013	(2.661, .486)	.065	(1.912, .999)	.392	(.866, .230)
8	21.97 (14.43)	.002	(3.429, .626)	.014	(2.614, 1.285)	.715	(.368, .093)
9	20.40 (17.48)	.002	(3.322, .607)	.012	(2.657, 1.571)	.881	(.150, .045)
10	20.37 (16.55)	.001	(3.520, .643)	.008	(2.812, 1.577)	.867	(.168, .048)

Note. JNW = just-noticeable-Whiteness. Δ represents Glass’s Δ , as variances between the PM and chain steps were unequal. Comparisons that survive Bonferroni correction at $p < .005$ are depicted in bold.

so as a case study in what such data look like (only for the FaceGen faces), and how they relate to the primary temporal analyses. In particular, we simply analyze reproduction biases without regard for which step in the chain a given participant was at. For example, these data for our FaceGen faces are depicted as a scatterplot in Figure 6, where each point represents the face that was reproduced (arrayed along the vertical axis) based on the face that was initially encountered (arrayed along the horizontal axis). As a result of this organization, points below the gray diagonal line represent faces that were biased Whiteward during reproduction, and points above the line represent faces biased Blackward. (Note that because each of the starting faces was in fact encountered during some particular chain, not all starting faces are represented equally—or, indeed, at all.)

Initial inspection of this scatterplot suggests two prominent patterns: (a) overall, there were just as many Blackward reproductions as there were Whiteward reproductions; but (b) those different categories did not occur equally in all regions of the face space: faces near the extremes were in general biased away from those extremes (perhaps in a sort of regression to the mean), but faces in the central (more ambiguous) region of the space were strongly biased Whiteward.

These impressions were verified via the statistical analyses reported below.

For the FaceGen faces, there were no more (Whiteward) faces below the line than (Blackward) faces above the line (31 vs. 25, $p = .50$ by an exact binomial sign test). However, if we focus on just those faces in the middle half of the spectrum (faces 16–46, which lie between the orange bars at the bottom of Figure 6—i.e., those that have the most space to deviate in either direction) we find that 22 of those 31 faces were biased Whiteward ($p = .03$). Further, a one-sample t test of the individual deviations from the 31 faces in that region revealed a robust Whiteward bias, relative to their corresponding starting faces ($MD = 3.63$, $SD = 11.97$), $t(175) = 4.02$, $p < .001$, $d = .30$. Thus, these faces in the central half of the space clearly replicate the Whiteward shift from the temporal analyses reported earlier.

In contrast, 13 of the 15 Whitest faces (faces 1–15, or the Whitest quarter of the faces—which lie to the left of the leftmost orange bar in Figure 6) were reproduced as Blacker ($p < .01$), which yielded a strong Blackward bias when the magnitudes of these deviations were considered ($MD = -5.23$, $SD = 10.11$), $t(103) = 5.28$, $p < .001$, $d =$

Table 3

Results of Temporal Analyses for Experiments 1–3, When Participants Viewed Luminance-Matched Averaged Photographs

Ss#	<i>M</i> (<i>SD</i>)	Versus middle		Versus perc. middle		Versus JNW	
		<i>p</i>	(<i>t</i> , <i>d</i>)	<i>p</i>	(<i>t</i> , Δ)	<i>p</i>	(<i>t</i> , Δ)
“Start Middle” chains (Experiment 1) – Average faces							
1	31.40 (13.98)	.877	(.157, .029)	.215	(1.265, .573)	<.001	(3.553, .794)
2	30.23 (15.44)	.788	(.272, .050)	.462	(.745, .371)	.006	(2.900, .703)
3	28.20 (14.36)	.294	(1.068, .195)	.968	(.041, .019)	.022	(2.379, .543)
4	25.07 (15.75)	.048	(2.063, .377)	.311	(1.031, .524)	.233	(1.210, .298)
5	22.67 (18.13)	.018	(2.517, .460)	.117	(1.614, .939)	.694	(.396, .110)
6	21.47 (16.85)	.004	(3.098, .566)	.042	(2.116, 1.147)	.951	(.062, .016)
7	23.70 (17.24)	.028	(2.319, .423)	.180	(1.372, .760)	.477	(.718, .191)
8	23.83 (15.69)	.018	(2.502, .457)	.155	(1.457, .737)	.417	(.821, .201)
9	25.13 (14.14)	.031	(2.272, .415)	.272	(1.117, .512)	.186	(1.344, .303)
10	25.97 (14.02)	.059	(1.966, .359)	.424	(.809, .368)	.107	(1.645, .369)
“Start Black-ish” chains (Experiment 2) – Average faces							
1	34.90 (10.03)	.042	(2.131, .389)	.001	(3.548, 1.179)	<.001	(6.111, 1.068)
2	30.57 (13.12)	.858	(.181, .033)	.322	(1.005, .429)	.001	(3.429, .729)
3	26.07 (16.20)	.106	(1.668, .305)	.507	(.671, .350)	.143	(1.492, .376)
4	26.53 (18.26)	.191	(1.340, .245)	.649	(.460, .270)	.148	(1.477, .413)
5	24.10 (15.99)	.025	(2.363, .431)	.190	(1.340, .691)	.378	(.891, .222)
6	19.23 (15.87)	<.001	(4.061, .741)	.005	(2.998, 1.534)	.526	(.640, .159)
7	21.23 (17.90)	.006	(2.988, .546)	.047	(2.066, 1.188)	.994	(.008, .002)
8	25.47 (16.78)	.081	(1.806, .330)	.407	(.841, .454)	.212	(1.267, .329)
9	24.07 (16.40)	.028	(2.315, .423)	.197	(1.319, .697)	.394	(.862, .220)
10	20.20 (15.94)	<.001	(3.711, .678)	.012	(2.659, 1.367)	.740	(.334, .083)
“Start White” chains (Experiment 3) – Average faces							
1	9.77 (9.78)	<.001	(11.893, 2.171)	<.001	(9.765, 3.174)	<.001	(5.236, .900)
2	11.97 (11.05)	<.001	(9.432, 1.722)	<.001	(7.681, 2.793)	<.001	(3.891, .728)
3	16.70 (14.94)	<.001	(5.244, .957)	<.001	(4.086, 1.973)	.137	(1.514, .357)
4	15.57 (15.69)	<.001	(5.387, .983)	<.001	(4.285, 2.169)	.077	(1.815, .446)
5	19.97 (14.15)	<.001	(4.272, .780)	.004	(3.070, 1.407)	.656	(.449, .101)
6	18.80 (13.01)	<.001	(5.138, .938)	<.001	(3.802, 1.609)	.366	(.912, .193)
7	19.90 (11.72)	<.001	(5.185, .947)	<.001	(3.694, 1.419)	.588	(.546, .106)
8	20.80 (13.36)	<.001	(4.181, .763)	.007	(2.908, 1.263)	.868	(.167, .036)
9	25.03 (16.79)	.061	(1.947, .355)	.335	(.980, .529)	.263	(1.136, .295)
10	24.33 (17.37)	.044	(2.102, .384)	.253	(1.165, .651)	.374	(.899, .241)

Note. JNW = just-noticeable-Whiteness. Δ represents Glass's Δ , as variances between the PM and chain steps were unequal. Comparisons that survive Bonferroni correction at $p < .005$ are depicted in bold.

.52. Similarly, eight of the 11 Blackest faces for which we had data (drawn from Faces 47–61, or the Blackest quarter of the faces—which lie to the right of the rightmost orange bar in Figure 6) were reproduced as Whiter ($p = .23$), which also yielded a very strong Whiteward bias when considering the deviation magnitudes ($MD = 11.35$, $SD = 13.47$), $t(19) = 3.77$, $p = .001$, $d = .84$. Both of these patterns could simply reflect a regression to the mean, from the extremes of the face space.

These atemporal analyses thus suggest that the robust Whiteward bias in the temporal analyses for FaceGen faces was driven largely by those faces in the central half of the space—such that the magnitude of the bias may be underestimated in the end (because a regression to the mean may effectively fight against the race-related bias as a face moves more and more Whiteward).

Experiment 2: Starting Black(-ish)

Does the Whiteward bias shown in Experiment 1 reflect convergence to a specific point in this face space, or does it just reflect a set amount of Whiteward drift over 10 serial reproductions? Here

we sought to test these possibilities by replicating the experiment, but this time starting each of the 30 ten-participant chains from a starting face that was approximately 75% of the way toward the Black extreme of the space.

Method

This experiment was identical to Experiment 1 except that each chain now began at face 46 (out of 61) instead of face 31 (i.e., the continuum's midpoint). We once again collected 30 chains of 10 White participants each for our two face continua, with exclusions detailed in Table 1.

Results and Discussion

The mean reproduction chains are depicted via the red lines in Figure 4 (FaceGen faces) and Figure 5 (averaged photographs). Inspection of these chains suggests three related patterns: (a) chains were again biased Whiteward, compared with the starting point; (b) there were more individual Whiteward steps before the

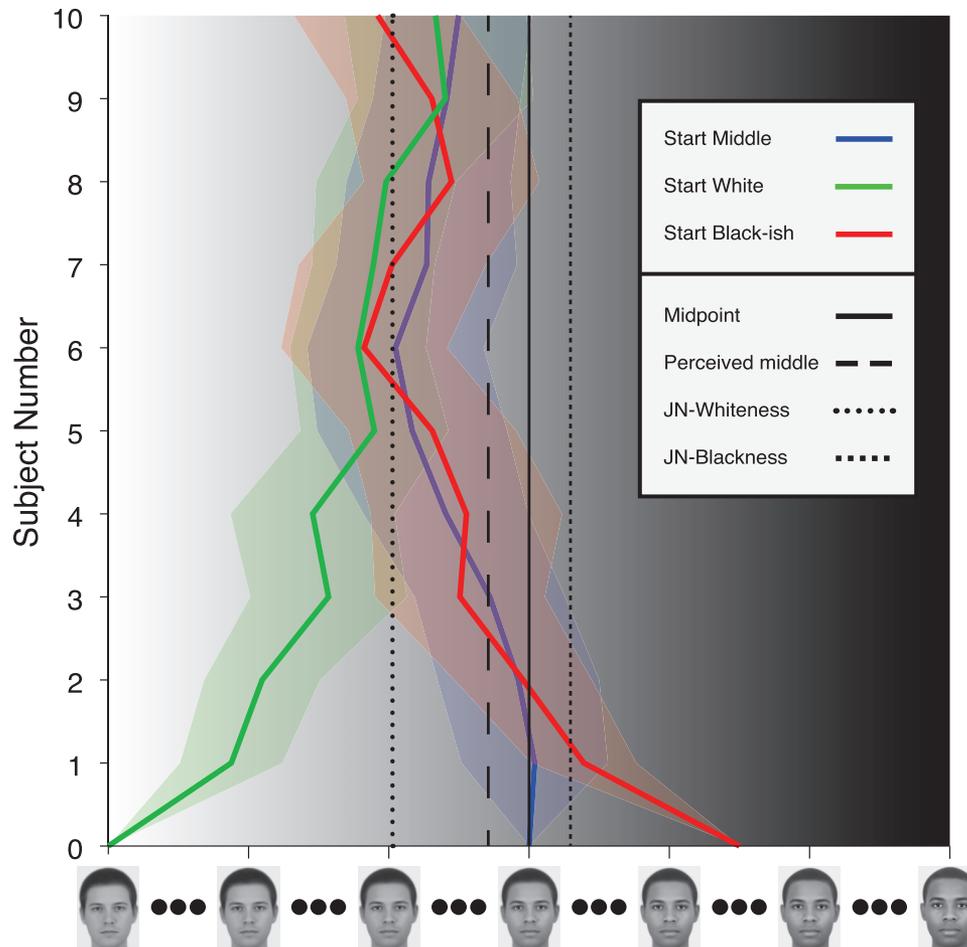


Figure 5. The mean chains of reproduction in a luminance-matched race continuum using averaged photographs. All data depicted are analogous to those shown in Figure 4. See the online article for the color version of this figure.

step-by-step bias diminished; and, correspondingly, (c) the magnitude of the overall bias over 10 reproductions was much greater than that in Experiment 1—effectively deviating by a greater amount (relative to the starting face), but to a similar absolute point well into the White region of the space. These impressions were verified via the statistical analyses reported below and in the middle section of Tables 2 and 3, all of which correspond to those described in Experiment 1.

For FaceGen faces (see Figure 4), 10 of the 10 points were again Whiteward of the starting Black-ish face, which of course is again a significant result by itself ($p = .002$). Furthermore, a series of t tests revealed that 9 of the 10 steps (#2–10) differed significantly from the starting point, even after correcting for multiple comparisons ($\alpha_{\text{corrected}} = .005$, all $ts[29] \geq 3.67$, $ps \leq .001$, $ds \geq .67$). These results indicate that there is a systematic and extremely strong Whiteward bias—such that the final point in the mean chain was itself more than 31 faces into White space relative to the starting point ($M = 14.93$, $SD = 9.92$), $t(29) = 17.15$, $p < .001$, $d = 3.13$ —or 16 faces into White space relative to the midpoint (roughly 54% of the way into White space). As detailed in the bottom row of the middle section of Table 2, this final step was

also significantly Whiteward of the continuum's midpoint. And now all 10 of the steps moved Whiteward relative to the previous step, as depicted visually in Figure 4 by the fact that every red line segment is sloped to the left ($p = .002$). Nevertheless, a series of Bonferroni-corrected paired-samples t tests revealed that only a single step (#3) differed significantly from its preceding step, ($\alpha_{\text{corrected}} = .005$, $t[29] = 3.34$, $p = .002$, $d = .61$; all other $ts[29] \leq 2.78$, all other $ps \geq .009$, all other $ds \leq .50$)—providing further evidence that individual jumps within the face space are relatively small even when they collectively move much farther, and that the bias Whiteward accrues with successive reproductions. The chain's ending point was thus not reliably different from the corresponding ending point in Experiment 1, $t(49.711) = 1.60$, $p = .12$, Glass's $\Delta = .348$, but the overall distance in face space that was traversed by the chain was much greater (31 faces vs. 10 faces), $t(49.711) = 6.10$, $p < .001$, Glass's $\Delta = 1.33$.

The results with averaged photographs (see Figure 5) were nearly identical. Ten out of the 10 points were again Whiteward of the starting Black-ish face ($p = .002$), and the t tests revealed that all 10 steps differed significantly from the starting point, even after correcting for multiple comparisons ($\alpha_{\text{corrected}} = .005$, all $ts[29] \geq$

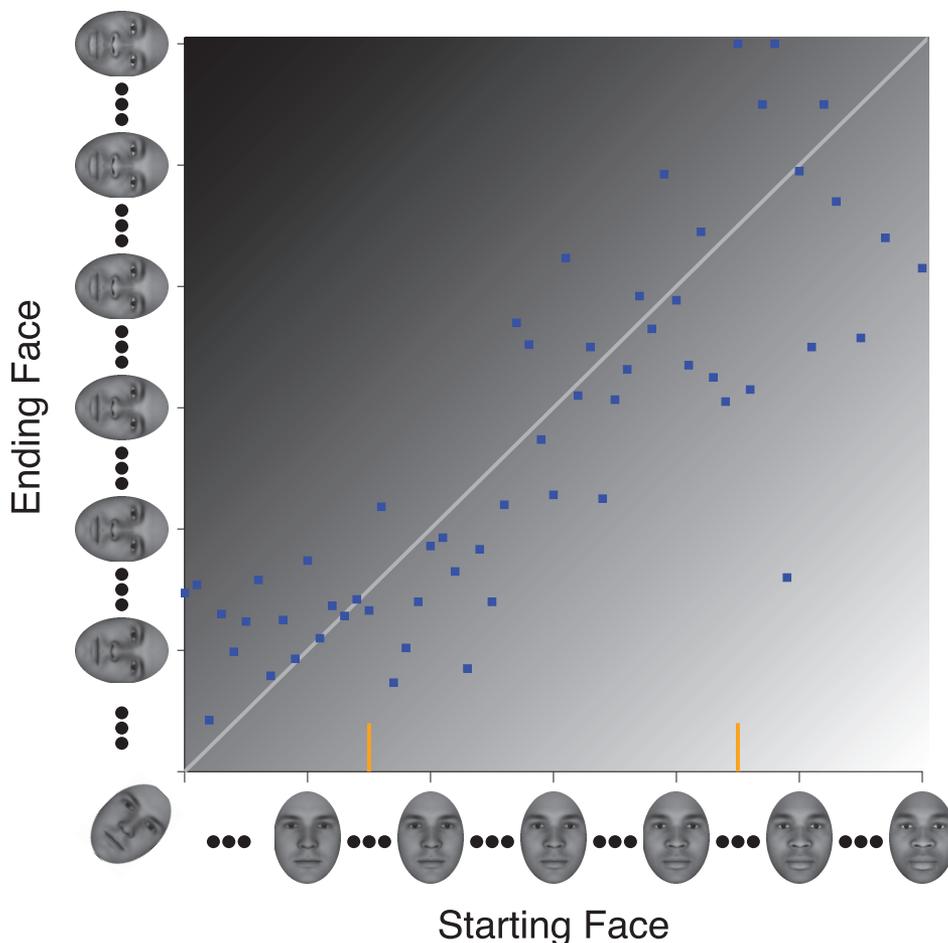


Figure 6. Atemporal depiction of White participants' response data with FaceGen faces from Experiment 1, plotting the mean outcome of participants' reproductions for each face shown to them over the course of the experiment. The horizontal axis depicts the face initially shown to participants, whereas the vertical axis represents the corresponding reproductions. The diagonal line thus represents perfect reproduction. Points below the line represent faces reproduced as Whiter than they were. Points above the line represent faces reproduced as Blacker than they were. The orange vertical lines indicate the bounds of the middle 50% of faces (see the text for details). See the online article for the color version of this figure.

5.83, $ps < .001$, $ds \geq 1.06$). These results again indicate a systematic and extremely strong Whiteward bias—such that the final point in the mean chain was itself more than 25 faces into White space relative to the starting point ($M = 20.20$, $SD = 15.94$), $t(29) = 8.87$, $p < .001$, $d = 1.62$ —or 11 faces into White space relative to the midpoint (roughly 35% of the way into White space). As detailed in the bottom row of the middle section of Table 3, this final step was also significantly Whiteward of the continuum's midpoint. And now seven of 10 steps moved Whiteward relative to the previous step, as depicted visually in Figure 5 by the leftward sloping red line segments ($p = .344$). Nevertheless, a series of Bonferroni-corrected paired-samples t tests revealed that no step differed significantly from its preceding step, ($\alpha_{\text{corrected}} = .005$, all $ts[29] \leq 2.27$, all $ps \geq .030$, all $ds \leq .42$). The average face chain's ending point was thus not reliably different from the corresponding ending point in Experiment 1, $t(58) = 1.49$, $p = .142$, $d = .39$, but the overall distance in face

space that was traversed by the chain was much greater (26 faces vs. five faces), $t(58) = 5.36$, $p < .001$, $d = 1.41$.

This pattern of results—as observed with both FaceGen faces and averaged photographs—suggests convergence toward a particular point in the face space rather than a set absolute average amount of Whiteward drift per step. It may be worth highlighting at this point how surprising these results were to us, for at least two reasons. First, the magnitude of the effect seemed especially striking in this study—with a starting face that was 50% of the way into the Black region ending up more than 50% of the way into the White region in only 10 steps. Second (and as explored in the General Discussion) these results seem like the opposite of those that one would expect from most previous studies of memory for race—most of which would suggest that memory would be distorted toward the nearest *extreme* of the continuum (e.g., Corneille, Huart, Becquart, & Brédart, 2004; Tanaka & Corneille, 2007).

Experiment 3: Starting (Maximally) White

Is the Whiteward bias observed in Experiments 1 and 2 driven only by movement toward the extreme White end of the space (suggesting that given a large enough number of subsequent reproductions, perhaps all mean chains would end up at the most White face)—or does it truly reflect convergence to a specific region of the space? Here we sought to test these possibilities by replicating Experiment 1, but this time starting each of the 30 ten-participant chains from a starting face that was maximally White (i.e., the face at the extreme White end of the space, depicted at the far left of Figure 2).

Method

This experiment was identical to Experiment 1 except that each chain now began at face 1 (out of 61) instead of face 31 (i.e., the continuum's midpoint).

Results and Discussion

The mean reproduction chains are depicted via the green lines in Figure 4 (FaceGen faces) and Figure 5 (averaged photographs). Inspection of these particular chains suggests three related patterns: (a) chains were now biased Blackward, compared with the (maximally White) starting point; (b) chains veered Blackward and away from the starting point rapidly (but incrementally) via a few small jumps and then proceeded to stay at a preferred point in the space for the rest of the chain (continuing in the graph as a largely vertical—albeit noisy—line, just as did the chains from Experiment 1, but in the opposite direction); and (c) the magnitude of the overall biases over 10 reproductions were intermediate (in the case of FaceGen faces) or similar (in the case of averaged photographs) to those observed in Experiments 1 and 2, which was a direct result of their convergence on a similar absolute point within the White region of the space. These impressions were verified via the statistical analyses reported below and in the bottom section of Tables 2 and 3, all of which correspond to those described in Experiments 1 and 2.

For FaceGen faces (see Figure 4), in contrast to our earlier experiments, 10 of the 10 points were now Blackward of the starting maximally White face ($p = .002$). Furthermore, a series of t tests revealed that all 10 steps differed significantly from the starting point, even after correcting for multiple comparisons ($\alpha_{\text{corrected}} = .005$, all $t_s[29] \geq 5.33$, $p_s < .001$, $d_s \geq .97$). These results indicate that there is a systematic Blackward bias—such that the final point in the mean chain was itself more than 19 faces away from the starting face ($M = 20.37$, $SD = 16.55$), $t(29) = 6.41$, $p < .001$, $d = 1.17$ —or 10 faces into White space relative to the midpoint (roughly 35% of the way into White space). As detailed in the bottom row of the bottom section of Table 2, this final step was also significantly Whiteward of the continuum's midpoint. However, only six of the 10 steps moved Blackward relative to the previous step, as depicted visually in Figure 4 by each green line segment which slopes to the right. Furthermore, a series of Bonferroni-corrected paired-samples t tests revealed that only a single step (#1) differed significantly from its preceding point, ($\alpha_{\text{corrected}} = .005$, $t[29] = 5.33$, $p < .001$, $d = .97$; all other $t_s[29] \leq 2.68$, all other $p_s \geq .01$, all other $d_s \leq .49$)—providing

yet further evidence that individual jumps within the face space are relatively small, and that the bias (Blackward, in this case) accrues with successive reproductions. Here, the chain's ending point was thus not reliably different from either the ending point of the chain from Experiment 1, $t(58) = .02$, $p = .98$, $d = .006$, or the ending point of the chain from Experiment 2, $t(47.470) = 1.54$, $p = .13$, Glass's $\Delta = .55$.

These analyses all yielded qualitatively similar results for averaged photographs (see Figure 5). Once again 10 of the 10 points were Blackward of the starting maximally White face ($p = .002$), and again t tests revealed that all 10 steps differed significantly from the starting point ($\alpha_{\text{corrected}} = .005$, all $t_s[29] \geq 4.91$, $p_s < .001$, $d_s \geq .89$). These results again show that there is a systematic Blackward bias—such that the final point in the mean chain was itself more than 23 faces away from the starting face ($M = 24.33$, $SD = 17.37$), $t(29) = 7.36$, $p < .001$, $d = 1.34$ —or six faces into White space relative to the midpoint (roughly 22% of the way into White space). As detailed in the bottom row of the bottom section of Table 3, this final step was also marginally significantly Whiteward of the continuum's midpoint. However, only seven of the 10 steps moved Blackward relative to the previous step, as depicted visually in Figure 5 by each rightward-sloping green line segment. Furthermore, a series of Bonferroni-corrected paired-samples t tests revealed that no step differed significantly from its preceding point, ($\alpha_{\text{corrected}} = .005$, all $t_s[29] \leq 2.16$, all $p_s \geq .039$, all $d_s \leq .40$). Here, the chain's ending point was thus not reliably different from either the ending point of the corresponding mean chain from Experiment 1, $t(58) = .40$, $p = .690$, $d = .11$, or the ending point of the chain from Experiment 2, $t(58) = .96$, $p = .341$, $d = .25$.

Across both face continua, we again observed the same pattern wherein the starting face gradually—but consistently—moved to a point well into the White region, suggesting convergence toward a region that is distinct from the extreme White end of the space.

Experiment 4: Perceived Middles?

The results of Experiments 1–3—corresponding to the three colored lines in Figures 4 and 5—clearly suggest a convergence to a particular point or region of this face space, but what is that point? It is clearly not the objective midpoint of the space (i.e., the most racially ambiguous face), because (per the results in Tables 2 and 3) each of these chains drifted significantly Whiteward of that midpoint. But could the point of convergence correspond instead to the *perceived* midpoint of the spectrum? This could be the case if, for example, racially ambiguous faces are perceived as being more akin to the minority group than to the majority group (as suggested in the so-called “one drop” rule, or the principle of hypodescent; e.g., Banks & Eberhardt, 1998; Halberstadt, Sherman, & Sherman, 2011; Peery & Bodenhausen, 2008). In this case, a face that was 50% White and 50% Black (i.e., the actual midpoint, depicted by the solid black vertical line in Figures 4 and 5) might be *perceived* as being more Black than White, whereas some face that was actually more White than Black would be perceived as 50% White. Could such a perceived midpoint correspond to the convergence observed in Experiments 1–3? To find out, we simply measured where that perceived midpoint was directly for each of our two face continua.

Method

The faces, recruitment details, attention check, and response measure (with a centrally presented face morphing from Black to White and back as participants moved their cursor around the surrounding ring) were identical to those used in Experiments 1–3.

Participants. We decided before data collection began to test 100 participants. This value was chosen arbitrarily to be roughly in line with previous large-sample studies run with this population, and the same sample size was used in the related Experiment 5.

Procedure. Participants first read detailed written instructions for the task, as described below. After clicking on a small box labeled “Start” (which was only available after all images to be shown had been preloaded), three stimuli appeared: the central response ring (now 380 px diameter), centered horizontally 37 px from the top of the browser window; and both the Whitest and Blackest faces (the same size as in Experiments 1–3, and dependent on the face continuum used), flanking (and centered vertically with) the central ring, separated by 64 px of blank space (one on the left and one on the right, chosen randomly for each observer). One second after these stimuli appeared, the same attention check as used in Experiments 1–3 was presented. Participants then moved their cursor around the central face while it morphed through the space (from Black to White and back again, as described below) and selected that face that they judged to be “the perfect blend” of the two flanking extremes. To ensure that this judgment had to be made on the basis of the visual appearance of the centrally morphing face itself (and not just on the position of the marker), the central face did not morph through the full 61 faces as it did in Experiments 1–3. Instead, the slider allowed the central face to morph continuously only through a restricted range of this space, with that range selected via a different random choice for each participant from two possibilities: (a) one of the end points of the morphing continuum was face 6 (i.e., five faces from the Whitest face), whereas the other end point was randomly chosen between faces 36 and 56; or (b) one of the end points was face 56 (i.e., five faces from the Blackest face), whereas the other end point was randomly chosen between faces 6 and 26. This procedure effectively unconfounded the middle of the full continuum between the visible flanking extremes and the middle of the available morphing continuum.

Once satisfied with their selection, participants clicked their mouse and pressed a box labeled ‘Next’ to record their response (as in Experiments 1–3), though now no response could be recorded until participants had moved the marker at least 30 different individual face morphs away from its random starting position. (Participants were instructed that “there is no time pressure” and that they should be “as accurate as you can in blending the faces.” Because of a coding error, however, participants were still excluded if they required more than 60s to complete this task in the case of the FaceGen faces—though (per Table 1) this only excluded eight participants who were then replaced.) Participants were then asked (via written prompt with attached response boxes) about (a) which countdown number had been red, (b) a description of the task they just completed, (c) whether they encountered any technical problems during the experiment, and (d) “What do you think we were testing in this survey?”

Results and Discussion

The mean responses are depicted as the black dashed vertical lines just to the left of the actual midpoint in Figure 4 (FaceGen faces) and Figure 5 (averaged photographs). Despite the fact that we randomly varied the range of faces each participant viewed, these responses were still highly consistent (FaceGen faces: $M = 29.00$, $SD = 5.47$; averaged photographs: $M = 28.09$, $SD = 5.77$). These “perceived middle” points in the continua were very close to the actual midpoints, though slightly Whiter (FaceGen faces: $t[99] = 3.65$, $p < .001$, $d = .37$; averaged photographs: $t[99] = 5.04$, $p < .001$, $d = .50$), and this was true for a majority of the participants (FaceGen faces: 66/100 participants, $p = .002$; averaged photographs: 69/100 participants, $p < .001$). However, as detailed in the middle columns of Tables 2 and 3, this perceived midpoint was significantly Blacker than each of the final steps from the chains of Experiments 1–3 for the FaceGen faces, but was not reliably Blacker than any of the final steps for the averaged photographs.

Thus, the point of convergence observed in the previous experiments was Whiter than the *perceived* midpoints of the continua (for the FaceGen faces) and Whiter than the *actual* midpoints of the continua (for both sets of faces). It is also telling that the perceived midpoints were significantly Whiter than the actual midpoints for both sets of face stimuli. And this may be for the same reason that memory appears biased toward the White end of the spectrum.

Experiments 5a and 5b: Just-Noticeable Whiteness and Blackness

Another possibility is that the point of convergence observed in Experiments 1–3 corresponds to the face that is “just noticeably White”—that is, to that face in the continuum that is perceived as just barely Whiter than the point at which a face is no longer seen as being racially ambiguous. To find out (and to confirm that the point of convergence is indeed noticeably White), we simply measured directly which face in the continuum was, on average, considered to be just-noticeably White (in Experiment 5a) and just noticeably Black (in Experiment 5b).

Method

These experiments were identical to Experiment 4 except as noted below.

Participants. We decided before data collection began to test 100 participants in each experiment, with this value chosen to match the sample size of Experiment 4, and with the exclusions detailed in Table 1.

Procedure. After clicking on the “Start” button, a single central face was presented (now without the surrounding ring or the flanking faces). In Experiment 5a, this face was always the Blackest face in the relevant face continuum. Participants then used their mouse to move a (20 px square) marker along a light gray horizontal slider (549 px by 14 px) that was centered beneath the face (separated by 202 px) with the marker always beginning at the leftmost position. As participants moved the marker along the slider, the face morphed between the Blackest face (at the leftmost starting position) to the Whitest face (at the rightmost position).

Participants were instructed to slowly move the slider to the right, and to stop as soon as they reached the very first face they identified as being White. Participants then clicked a “Next” button (centered below the slider) to continue (though they were allowed to reposition the marker first if they overshot). Experiment 5b was identical, except that the Whitest face was initially presented, linked to the leftmost slider position (with the Blackest face then linked to the rightmost slider position), and participants were instructed to stop as soon as they reached the very first face they identified as being Black. Participants then answered a simple arithmetic problem (such as “What is eight plus seven?” or “What is four plus three?”) as an attention check.

Results and Discussion

The mean responses from Experiment 5a (just-noticeable-Whiteness, or JNW) are depicted by the black dotted vertical lines to the left of the midpoint in Figure 4 (FaceGen faces; $M = 20.91$, $SD = 11.42$) and Figure 5 (averaged photographs; $M = 21.26$, $SD = 12.77$). And the mean responses from Experiment 5b (just-noticeable-Blackness, or JNB) are depicted by the black dotted vertical lines to the right of the midpoint in Figure 4 (FaceGen faces; $M = 35.67$, $SD = 10.83$) and Figure 5 (averaged photographs; $M = 33.95$, $SD = 11.94$). As expected, the JNB was Blacker than the midpoint for both face continua, $t(99) = 4.31$, $p < .001$, $d = .43$; $t(99) = 2.47$, $p = .015$, $d = .25$. And (because the point of convergence observed previously was always in White space) this JNB was Blacker than the final step of the mean chains in each of Experiments 1–3 for both face continua (all $t_s \geq 2.84$, all $p_s < .008$, all $Glass's \Delta_s \geq .66$). Also as expected, the JNW was Whiter than the midpoint for both face sets, $t(99) = 8.83$, $p < .001$, $d = .88$; $t(99) = 7.63$, $p < .001$, $d = .76$. Most critically, however, and as detailed in the rightmost columns of Table 2, the JNW for FaceGen faces was *not* reliably different than either of the final steps from the mean chains of Experiments 1 or 3 (which were less than one face away)—though it was Blacker than the final step from the corresponding mean chain of Experiment 2 (by about six faces). Similarly, as detailed in the rightmost columns of Table 3, the JNW for averaged photographs was not reliably different than any of the final steps from the corresponding mean chains of Experiments 1–3. Thus, the point of convergence observed in Experiments 1–3 was always *at least as White* as the JNW—confirming that the point of convergence is recognizably White for both face continua.

Experiments 6a and 6b: Any Old Continuum? (Morphing From Bob to Dan)

We have been assuming so far that the point of convergence revealed in Experiments 1–3 corresponds to some region of the particular face space we have been using, per se. But it is also possible that this sort of convergence is driven not by the faces (or indeed, by any aspect of the stimuli) but rather by the response measure itself. Perhaps, for example, when choosing a particular response from a scale under some uncertainty, participants are simply inclined to choose a seemingly arbitrary point—and so they stay away from seemingly *non*-arbitrary points such as the end points or the midpoint, as observed during the actual morphing in the response phase. This would yield mean responses somewhere

between the midpoint and extremes—similar to either the point of convergence observed in Experiments 1–3, or its mirror image on the other side of the continuum. As an alternate explanation for our results, this possibility seems woefully inadequate—because it cannot (by definition) explain the central trend in those results, which is that all of the chains converged on a preferred region well into (only) *White* space. Nevertheless, to highlight that the fact that the results of Experiments 1–3 reflected something particular to *race* (i.e., to the stimuli, and not just to the abstract response scale, the task mechanics, or generic response strategies), we effectively replicated Experiments 1 and 2 but now using a face continuum constructed from a pair of arbitrary exemplar face identities (‘Bob’ and ‘Dan’), as depicted in the horizontal axis labels of Figure 7.

Method

These experiments were identical to Experiments 1 and 2 except as follows.

Stimuli. Faces (241 px by 238 px) were generated in a similar fashion to those of the FaceGen faces from the previous experiments. We generated two random White 30-year-old male anchor faces (‘Bob’ and ‘Dan’) using the “European” ethnicity tab and the “Generate” function in FaceGen Modeler 3.5 (Singular Inversions, <http://www.facegen.com>), and created a smooth morph continuum of 61 faces (all full-color and uncropped) using FantaMorph (Abrosoft, <http://www.abrosoft.com>).

Procedure. Whereas the chains collected in Experiment 1 began with the midpoint between Black and White, the chains collected in Experiment 6a began with the midpoint between Bob and Dan. And whereas the chains collected in Experiment 2 began with a “Black-ish” face, the chains collected in Experiment 6b began with the equivalent “Dan-ish” face.

Results

The mean reproduction chains of Experiments 6a and 6b are depicted via the blue line and the red line, respectively, in Figure 7. Inspection of this figure reveals a clear pattern: by the first step, *both* mean chains had simply converged on the midpoint of the space, after which they continued as noisy vertical lines. These impressions were confirmed via the statistical analyses reported below and in Table 4.

In Experiment 6a, as detailed in Table 4, a series of Bonferroni-corrected t tests revealed that none of the steps differed from the starting (mid-)point. These results indicate that there was no systematic (Dan-ward or Bob-ward) bias. Accordingly, only six of the 10 steps moved in the leftward (i.e., Bob-ward) direction relative to the previous step, as depicted visually in Figure 7 by the number of times the blue line slopes to the left versus right. Moreover, a series of Bonferroni-corrected t tests confirmed that no step differed significantly from its preceding step, ($\alpha_{\text{corrected}} = .005$, all $t_s \leq 1.72$; all $p_s \geq .09$; all $d_s \leq .32$).

In Experiment 6b, a series of Bonferroni-corrected t tests revealed that all of the steps differed from the starting (Dan-ish) point (all $t_s \geq 3.31$; all $p_s \leq .003$; all $d_s \geq .60$), and the fact that all 10 of these steps were leftward (or “Bob-ward”) of the starting face is itself a significant result ($p = .002$). However, as detailed in Table 4, none of these steps differed from the midpoint. Seven of the 10 steps moved in the leftward (i.e., Bob-ward) direction

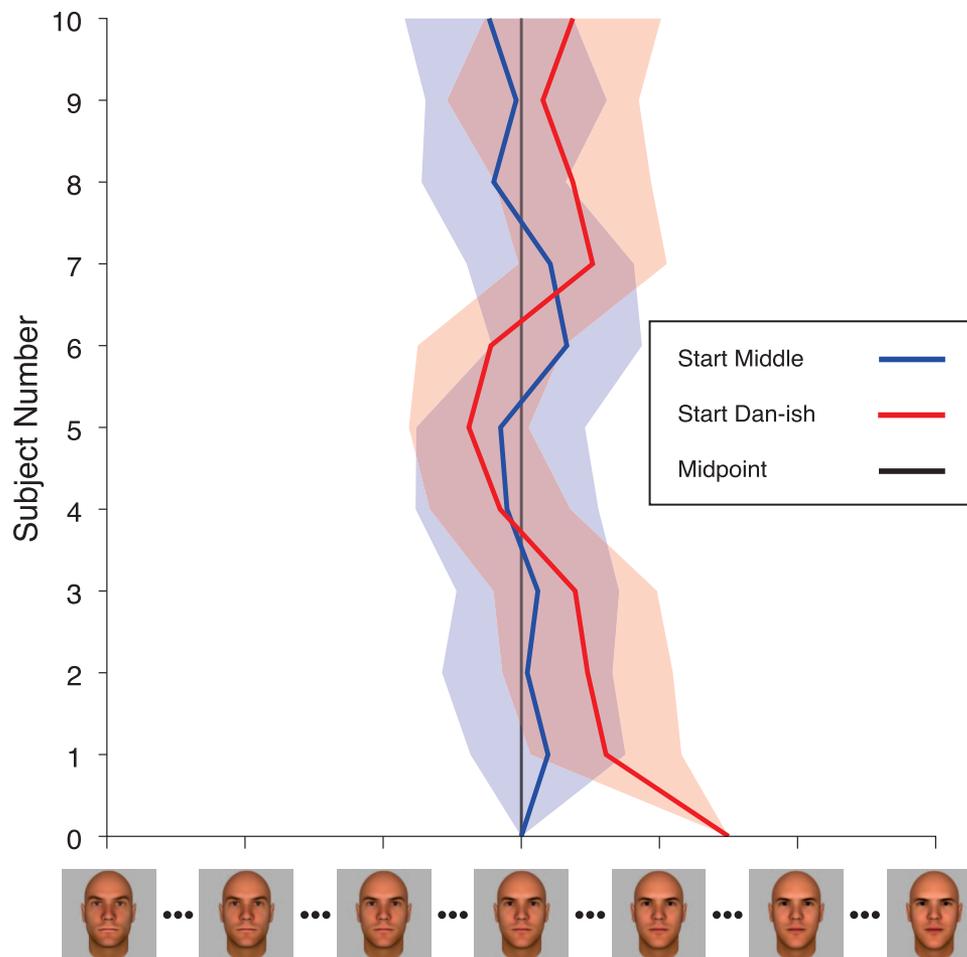


Figure 7. Mean chains of reproduction for Experiment 6a and 6b, in the same format as Figures 4 and 5. Moving from left to right along the horizontal axis takes one from the 'Bob' end to the 'Dan' end of the space. See the online article for the color version of this figure.

relative to the previous step, as depicted visually in Figure 7 by the number of times the red line slopes to the left. Moreover, a series of Bonferroni-corrected t tests revealed that only the first step differed significantly from its preceding step, ($\alpha_{\text{corrected}} = .005$, $t[29] = 3.32$, $p = .002$, $d = .61$). The chain's ending point was thus not reliably different from the ending point in Experiment 6a, $t(58) = 1.40$, $p = .17$, $d = .37$, but the overall distance in face space that was traversed by the chain was greater (two faces vs. 11 faces), $t(58) = 2.07$, $p = .04$, $d = .54$.

These results were notably different than those from the earlier experiments: whereas the race-related chains (in Experiments 1–3) converged on a region that was well into White space, the chains from the current Bob/Dan experiments simply converged on the midpoint. This is clear from the stark difference between Figures 4/5 and 7, and can be shown most clearly by comparing the similarly colored lines in these two graphs. For example, the final step of the mean FaceGen chain in Experiment 1 (Start Middle with White/Black) is significantly to the left of the final step in the mean chain in Experiment 6a (Start Middle with Bob/Dan), $t(58) = 2.05$, $p = .04$, $d = .54$, though this was not the case for the averaged photographs, $t(58) = 1.62$, $p = .111$, $d = .42$. However,

the final step of the mean FaceGen chain in Experiment 2 (Start Black-ish) is also significantly to the left of the final step in the mean chain in Experiment 6b (Start Dan-ish), $t(46.547) = 5.49$, $p < .001$, Glass's $\Delta = 2.00$, as is the final step of the mean chain for averaged photographs, $t(58) = 3.41$, $p = .001$, $d = .89$.

Discussion

It is unclear why the chains in these experiments converged on the midpoint between Bob and Dan. Perhaps this just reflects some generic attraction to the midpoint of any arbitrary scale in this context. (If so, that would suggest that the race-related point of convergence in Experiments 1–3 is actually *further* into White space than is suggested by those results, if in fact such an attractor is effectively competing with a sort of 'midpoint bias'.)

In any case, the critical aspect of these results is how dramatically they differed from those of Experiments 1–3: whereas the point of convergence for race was to a particular left-of-center region in the face space (as in Figure 1d), the Bob/Dan chains in these experiments simply converged on the midpoint (as in Figure 1c). This confirms that the chains from Experiments 1–3 must

Table 4
Results of Temporal Analyses for Experiments 6a and 6b

Ss#	Versus middle		(t, d)
	M (SD)	p	
"Start Middle" chains (Experiment 6a)			
1	32.93 (15.03)	.487	(.704, .129)
2	31.43 (16.53)	.887	(.144, .026)
3	32.20 (15.78)	.680	(.417, .076)
4	29.97 (17.72)	.752	(.319, .058)
5	29.50 (16.35)	.619	(.502, .092)
6	34.30 (14.55)	.224	(1.242, .227)
7	33.10 (16.21)	.484	(.709, .130)
8	29.00 (13.98)	.440	(.784, .143)
9	30.63 (17.58)	.910	(.114, .021)
10	28.67 (16.36)	.441	(.781, .143)
"Start Dan-ish" chains (Experiment 6b)			
1	37.13 (14.63)	.029	(2.296, .419)
2	35.80 (16.54)	.123	(1.590, .290)
3	34.90 (15.86)	.188	(1.347, .246)
4	29.47 (13.64)	.543	(.616, .112)
5	27.20 (11.61)	.084	(1.792, .327)
6	28.80 (14.18)	.402	(.850, .155)
7	36.17 (14.39)	.059	(1.967, .359)
8	34.73 (15.16)	.188	(1.349, .246)
9	32.57 (18.59)	.648	(.461, .084)
10	34.73 (17.10)	.241	(1.196, .218)

Note. No comparisons are significant, even at $p < .05$ (uncorrected for multiple comparisons).

reflect something particular to *race*, and not just to any generic (i.e., stimulus-independent) strategies or task demands.

General Discussion

The eight experiments reported here collectively demonstrate a new type of race-related bias in face memory. This bias was revealed by exploring memory with the method of serial reproduction. Other recent uses of this method have been argued to reveal types of 'mental defaults,' as the chains of reproduction, faced with some noise or uncertainty, converge on what the observers' minds are collectively biased to infer the 'message' is most likely to be (e.g., Bangert, 2000; Bartlett, 1932; Kalish et al., 2007; Lyons & Kashima, 2001; Mesoudi et al., 2006; Xu & Griffiths, 2010).

As applied to faces in our 'TeleFace' variant, we observed a similar 'default' in that the reproduction chains reliably converged on particular regions of the face space. In particular, focusing on White participants' reproductions of particular faces drawn from a continuum from White to Black (while equating mean luminance), the results of these experiments (as depicted in Figures 4 and 5) revealed chains that were consistently biased to a region well within White space—regardless of whether those chains started elsewhere in White space (in Experiment 3), at the midpoint of the continuum (in Experiment 1), or even well into Black space (in Experiment 2). This pattern of convergence did not reflect any sorts of generic strategies or methodological factors, since very different sorts of chains resulted when the same method was applied to an arbitrary face continuum (in Experiment 6). And the particular White-ish bias we observed was not equivalent to the

midpoint of the continuum (Experiments 1–3). Instead, it may be better explained by appeal to the *perceived* midpoint of the continuum (at least for averaged photographs; Experiment 4), or perhaps the face that is just barely recognizable as unambiguously White (Experiment 5).

Previous work suggests that such patterns of convergence in serial reproduction persist even when the chains continue past the initial stages of this convergence, such that adding additional participants to the chains does not qualitatively affect the outcome (e.g., Griffiths et al., 2006; Kalish et al., 2007; Xu & Griffiths, 2010). And this can be seen in the present work, which tested 10-step chains even though the convergence appeared to have been largely completed by around the fifth step. This can be observed in Figures 4 and 5, because the variously colored lines trace unique paths until approximately the fifth step, at which point they start to overlap along noisy mostly vertical trajectories. This suggests that our chain lengths were easily sufficient to observe the underlying patterns of convergence.

What Is the Nature of This Whiteward Bias?

It is perhaps worth emphasizing that the particular sort of convergence we observed in these experiments was not a foregone conclusion. Indeed, before running these studies, we did not have any clear hypothesis about how the chains would unfold—and if anything, we thought that a pattern such as Figure 1f was most likely, where any small differences would be amplified over time, essentially making the stimuli less and less racially ambiguous as the chains proceeded. Indeed, that is precisely what one would have to expect based on previous studies of biases in memory for faces—which has generally shown memory distortion toward more prototypical category exemplars (albeit with much larger distances between memory foils, compared with the tiny steps between adjacent faces in the test morphs; e.g., Corneille et al., 2004; Halberstadt & Niedenthal, 2001; Huat, Corneille, & Becquart, 2005). As this is clearly not what occurs, we may then inquire as to what the actual convergence (as in Figures 4 and 5) *does* represent. Of course, further research will be required to answer this question, but there are several possibilities.

Perhaps the most salient possible explanation of our results is that memory for race is being biased toward a statistical average—equivalent not to the midpoint of the relevant continuum, but rather to the *average point* along that continuum, as a function of all of the individual faces that one has actually encountered. Such a possibility might apply in this case, since our participants were all White U.S. residents—and U.S. demographics are such that if one happens to see a random face divorced from any context, it is many times more likely to be the face of a White person than a Black or mixed-race one (Humes, Jones, & Ramirez, 2011).

There are other salient possibilities as well, however. Perhaps, for example, the convergence is driven not by the average of all the faces one has experienced over an entire life, but is instead a function of the average of the faces seen only *recently*—which might be more adaptive, so as not to place undue weight on statistically 'obsolete' information. Or, in direct contrast, perhaps the bias is based on only the statistical properties of those faces encountered early on in one's life—perhaps during some critical period in which a default or prior was first formed (as with accents in speech). (We were particularly intrigued to see the recent report

that certain abilities related to face perception are correlated with the size of one's hometown; Balas & Saville, 2015.)

Or, moving to an entirely different form of explanation, it is also possible that the convergence we observed relates not to one's experience with other faces, but rather to one's *own* identity. Recall that all of our participants in the serial 'race' reproduction studies were White. Because of this, it is possible that this population's region of convergence was essentially *themselves*.

At least some of these views make clearly different predictions that could be tested in future work along these lines. Consider, to take one example, a population of Black participants who live (and grew up) in an environment populated mostly by White people. If the convergence is fueled by the statistical mean of one's experiences, then this population might effectively replicate the results of Experiments 1–3, producing chains that are qualitatively similar to those in Figures 4 and 5. But if the convergence is fueled in some way by one's own identity, then this population may produce chains that are in effect the *opposite* from the ones we observed, as if the lines in Figures 4 and 5 were mirror-reversed while keeping the axes the same.

More generally, it will be important for future work to determine just how specific our observed results are to the White U.S.-based MTurk population. On one hand, we chose this platform precisely because of its ability to reach a broad, general population. And in this regard, MTurk is surely far more diverse—and representative of the broader U.S. population—than is, say, a sample of college students (see Levay, Freese, & Druckman, 2016). And the robustness of the observed results also places some constraints on how much different subgroups in this population can differ from each other—since if any given participant in a chain has a highly *different* “prior” than the preceding participant in that chain, then (a) the chains will simply never converge on any particular region, and/or (b) the error bars around those chains will never be as narrow as we observed (as would be the case, e.g., if the observed point of convergence was an average of some highly variable bimodal distribution of responses across chains). The data thus suggest that there is a strong general trend/prior in the population as a whole—which of course is what our study aimed to characterize. On the other hand, though, there may be some subtler differences among subgroups (perhaps related to factors such as age, political orientation, socioeconomic status, level of media exposure, et al.) which future work could tease out—and of course this population is more uniform in terms of factors such as technological savviness (though we know of no work linking such factors to face or race perception).

Beyond such questions about possible subgroups, the magnitude (and perhaps even the direction?) of the memory biases we observed could also be modulated by various psychological factors, at an individual level—such as implicit racial bias (e.g., Walker & Hewstone, 2008). Or, to take another example, perhaps participants' motivations when doing the task (per social-cognitive theories of race processing) could influence how faces are remembered (Hugenberg et al., 2010; Young et al., 2012). For example, explicitly *labeling* our initially presented racially ambiguous faces as either “White” or “Black” could lead to subsequent memory biases toward the relevant category label (e.g., Hourihan et al., 2013; Pauker et al., 2009). We could not observe such effects in the present studies, however, since they were designed precisely to avoid any such task demands.

Contrasts With Categorical Memory Biases and the “One-Drop” Rule

The Whiteward bias observed in our studies is clearly *not* attributable to a type of categorical bias that has been observed in some previous (nonserial) studies of memory for faces. In the literature on racial categorization, for example, one of the core insights is the principle of hypodescent, or the “one-drop rule.” This is the phenomenon whereby a mixed-race person—say, one who is half-Black and half-White—is categorized as a member of the minority group by default (Banks & Eberhardt, 1998). This sort of principle, beyond its status as a psychological phenomenon, has also manifested itself in ugly historical trends. A one-drop rule was enforced during the history of the United States to disenfranchise people, for example—as when “one drop of Black blood makes a person Black” (Hickman, 1997, p. 1163) for the purposes of everything from census-taking to legal rights (Banks & Eberhardt, 1998; Hickman, 1997). (In the United States today, this principle is often noted when discussing Barack Obama, who is widely hailed as the “the first Black President,” despite having had a White American mother and a Black Kenyan father.)

It has been argued that the one-drop rule operates in part because of reflexive cognitive processes of attention and categorization, whereby racially ambiguous faces are automatically categorized as members of the minority group—especially under load and time pressure, and perhaps because of automatic aspects of attentional selection (Halberstadt et al., 2011; Peery & Bodenhausen, 2008). These categorization results are in effect the *opposite* of the pattern we observed with serial reproduction: whereas a principle of hypodescent must predict that one of our racially ambiguous faces would be categorized as Blacker than it really was, our results (i.e., the blue and red lines in Figures 4 and 5) clearly demonstrate such faces being reproduced as *Whiter* than they really were.⁷ Similarly, our findings are directly opposite to the patterns that are observed in at least some simpler studies of memory for faces: whereas we observed a region of convergence well within White space even for chains that began with Black faces (in Experiment 2, corresponding to the red lines in Figures 4 and 5), other (nonserial) studies of memory demonstrate that Black faces are in general likely to be remembered as *Blacker* than they really were (Corneille et al., 2004). These contrasts highlight the fact that the convergence we observed through serial reproduction might be strongly related to some properties of face perception

⁷ Note, however, that our results did yield subtler evidence of the principle of hypodescent in another way, when they measured the point at which the continuum from Black to White became unambiguously White (in Experiment 5a), and the point at which the continuum from White to Black became unambiguously Black (in Experiment 5b). Note that these faces—depicted in Figures 4 and 5 by the dotted vertical black lines—were not equally spaced around the objective midpoint of the continuum. Instead, a face effectively needed to be less Black (relative to the midpoint) to be judged as unambiguously Black, than it needed to be White (relative to the midpoint) to be judged as unambiguously White. (In Figures 4 and 5, this is visible via the fact that the dotted lines to the right of the midpoint are much closer to the midpoint than are the dotted lines to the left of the midpoint.) This asymmetry may even be for the same reasons that the perceived middles of the continua are to the left of their respective midpoints, and that the mean reproduction chains converge in White space. Perhaps the point of convergence draws both memory and category boundaries toward itself, as a function of one's visual experience with faces.

(such as the norm in norm-based coding models; Rhodes & Jeffery, 2006), but directly counter to others (such as categorical biases in perception and memory for faces).

Broader Societal Connections

Though the White default we observed may run counter to some other sorts of memory biases that have been observed for faces in the literatures on memory and categorization, such a default is of course not unknown in broader society, especially within the United States. For example, from the semiotics literature we know that Whiteness is usually ‘unmarked’ and is simply assumed to be the racial identity of a given person unless otherwise specified (e.g., Waugh, 1982). Similarly, within the social cognition literature it has been argued that there is a ‘White male default,’ in which persons are rarely categorized on the basis of their race (or sex) if they are White (or male), but are categorized on those bases when they deviate from those categories (e.g., Zárate & Smith, 1990). This realization has even trickled into popular media, as when the comedic TV host Stephen Colbert noted that “White male is American neutral,” in response to a controversy over racial diversity among smiley-face (“emoji”) icons. However, to our knowledge the serial reproduction results with White observers that we have reported here constitute the first time that a White “default” has been demonstrated within face perception.

Why Serial Reproduction?

We opted to use the method of serial reproduction in this study rather than other (perhaps more familiar) methods that have been used to study memory biases. For example, rather than testing only those faces ‘produced’ by other (previous-in-the-chain) observers, we could instead have systematically tested participants’ memory biases for every face in our continua—and then we could perhaps have indirectly ‘recovered’ the patterns of convergence by mapping out particular pseudochains after the fact (e.g., noting that Face 31 was misremembered on average as Face 24, which was in turn misremembered on average as Face 10, etc.).

As such, we are hesitant to suggest that the ‘TeleFace’ approach used here has any substantive advantages over other sorts of memory tests. Nevertheless, we think that this particular ‘stylistic variant’ of memory testing has at least three advantages, in terms of how it may answer certain questions related to memory in an especially direct way:

First, whereas the Whiteward bias could in principle be recovered from the ‘statistical wash’ of any sort of memory testing, it is especially directly and intuitively ‘read out’ from the patterns of convergence during serial reproduction—as is hopefully immediately appreciable in Figures 4 and 5.

Second, using serial reproduction allowed us to focus our key analyses on only a subset of points along the continuum—in particular, focusing on the average ending points of the chains, and how those points compare to both each other and to a few other salient points (e.g., the perceived middle of the continuum). In contrast, if we had instead explicitly mapped out responses to all (or even many) of the faces in the continuum, we would have ended up with a vast number of comparisons to test—which would both raise the specter of multiple comparisons and would effectively bury the data that *matter* for questions about shared inductive biases in a sea of other less relevant comparisons.

Third, serial reproduction may be better able to ‘scale up’ to questions about larger and more multidimensional spaces. Even staying in the realm of face spaces, for example, we might want to explore memory biases in face spaces that vary in both race and gender, where this would amount to a two-dimensional face space of 3,721 cells (whereas each dimension on its own would have only 61 cells, as in the horizontal axes of Figures 4 and 5). And whereas the method of serial reproduction could be applied just as readily to such multidimensional spaces (where the chains would now be ‘worms’ that could move through the greater-dimensional space), the test-the-entire-space approach would become practically impossible. (Even just attempting to Test 10 independent observations for each cell in such a space would require more than 35,000 subjects, which is vastly more than the average number of workers that Mechanical Turk has available for such experiments at any given time; see Stewart et al., 2015.) All of these reasons suggest that serial reproduction might be an especially useful approach to questions about memory biases in perceptual spaces.

Note that all of these advantages of serial reproduction hold even in cases (such as the present article) where the particular information that is being serially relayed is highly unusual from an ecological perspective. Although people actually do actually transmit stories to each other serially (both during games and during everyday social life), this essentially never occurs with faces (except perhaps in unusual circumstances such as those that involve police sketches).⁸ But our motivation in this work is manifestly not to explore some common social activity, but rather to explore a particular question psychological question using the method that seems most powerful and intuitive. In this way, we are hoping in this work to bring the notable success that serial reproduction has enjoyed in other domains into the study of face perception (and so to visual representations more generally). Just as with faces, people also do not typically ‘whisper’ mathematical functions or spatial layouts or estimates of causal strength to each other—and yet the use of serial reproduction in these domains (see Kalish et al., 2007; Lew & Vul, 2015; Yeung & Griffiths, 2015) has led to great insights anyway, because of its ability to reveal shared inductive biases in such a direct and intuitive way.

A General Tool?

Though the method of serial reproduction has a storied history in cognitive psychology and memory research, ever since Bartlett, it is used in contemporary research much more often to study factors related to culture (e.g., Bangerter, 2000; Lyons & Kashima, 2001; Mesoudi et al., 2006) than to study aspects of visual representation. And when it has been used in the context of iterated learning in recent work, it has sometimes been discussed more in terms of learning abstract properties and functions than visual representations themselves (e.g., Kalish et al., 2007). Nevertheless, and we hope as is illustrated by the present results, we think that this method may be a useful way to test memory when applied to questions about visual representation.

⁸ In light of our findings, future work should explore whether such police sketches may be biased toward our ‘mental defaults’ for faces. This should be especially salient along the Black-White dimension of race tested in our experiments.

Within the domain of face perception, for example, there is no reason why such studies are limited to race. Indeed, it seems possible in principle to explore biases using serial reproduction for any dimension of continuous face spaces—from gender and age to trustworthiness and attractiveness. And of course there is no reason in principle why such studies must explore unidimensional face spaces. For example, one could instead explore a two-dimensional space of race and gender—which might be a fascinating way of studying the social psychological notion of “gendered races” (Johnson, Freeman, & Pauker, 2012). One could even explore the influence of contextual factors on the patterns of convergence—as when stereotypical types of clothing related to status cues may influence the perception of racially ambiguous faces (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011). (And, of course, there is also nothing face-specific about this method, which could be used in principle to study many other types of visual representation.)

Conclusion

Face spaces, beyond serving as a framework for representing faces in general, may include certain prioritized regions. Such regions may serve not only as potential landmarks relative to which other faces may be coded, but may also bias downstream perception and memory. The present results, as a case study, highlight a clear Whiteward bias in memory for White observers, along the (luminance-matched) continuum between White and Black faces—and they emphasize the previously untapped potential of the ‘TeleFace’ method for directly revealing such biases.

References

- Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., . . . Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science, 17*, 836–840. <http://dx.doi.org/10.1111/j.1467-9280.2006.01790.x>
- Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin, 18*, 296–301. <http://dx.doi.org/10.1177/0146167292183005>
- Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia, 69*, 211–217. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.02.005>
- Bangerter, A. (2000). Transformation between scientific and social representations of conception: The method of serial reproduction. *British Journal of Social Psychology, 39*, 521–535. <http://dx.doi.org/10.1348/014466600164615>
- Banks, R., & Eberhardt, J. (1998). Social psychological processes and the legal bases of racial categorization. In J. Eberhardt & S. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 54–75). Thousand Oaks, CA: Sage.
- Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.
- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In W. Waggenspack (Ed.), *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 187–194). New York, NY: ACM Press.
- Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15*, 19–25. <http://dx.doi.org/10.1177/0146167289151002>
- Brooks, K. R., & Gwinn, O. S. (2010). No role for lightness in the perception of black and white? Simultaneous contrast affects perceived skin tone, but not perceived race. *Perception, 39*, 1142–1145. <http://dx.doi.org/10.1068/p6703>
- Canini, K. R., Griffiths, T. L., Vanpaemel, W., & Kalish, M. L. (2014). Revealing human inductive biases for category learning by simulating cultural transmission. *Psychonomic Bulletin & Review, 21*, 785–793. <http://dx.doi.org/10.3758/s13423-013-0556-3>
- Carbon, C.-C., & Albrecht, S. (2012). Bartlett's schema theory: The unreplicated “portrait d'homme” series from 1932. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 65*, 2258–2270. <http://dx.doi.org/10.1080/17470218.2012.696121>
- Chang, L., & Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell, 169*, 1013–1028.e14. <http://dx.doi.org/10.1016/j.cell.2017.05.011>
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 48*, 879–894. <http://dx.doi.org/10.1080/14640749508401421>
- Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts toward more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology, 86*, 236–250. <http://dx.doi.org/10.1037/0022-3514.86.2.236>
- Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7*, 173–179. [http://dx.doi.org/10.1016/S1364-6613\(03\)00057-3](http://dx.doi.org/10.1016/S1364-6613(03)00057-3)
- Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE, 8*, e57410. <http://dx.doi.org/10.1371/journal.pone.0057410>
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE, 6*, e25107. <http://dx.doi.org/10.1371/journal.pone.0025107>
- Germiné, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*, 847–857. <http://dx.doi.org/10.3758/s13423-012-0296-9>
- Goodman, N. D., Frank, M. C., Griffiths, T. L., Tenenbaum, J. B., Battaglia, P. W., & Hamrick, J. B. (2015). Relevant and robust: A response to Marcus and Davis (2013). *Psychological Science, 26*, 539–541. <http://dx.doi.org/10.1177/0956797614559544>
- Griffiths, T. L., Christian, B. R., & Kalish, M. L. (2006). Revealing priors on category structures through iterated learning. In *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 1394–1399). Mahwah, NJ: Erlbaum.
- Gwinn, O. S., & Brooks, K. R. (2015). No role for lightness in the encoding of Black and White: Race-contingent aftereffects depend on facial morphology, not facial luminance. *Visual Cognition, 23*, 597–611. <http://dx.doi.org/10.1080/13506285.2015.1061085>
- Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology, 81*, 587–598. <http://dx.doi.org/10.1037/0022-3514.81.4.587>
- Halberstadt, J., Sherman, S. J., & Sherman, J. W. (2011). Why Barack Obama is black: A cognitive account of hypodescent. *Psychological Science, 22*, 29–33. <http://dx.doi.org/10.1177/0956797610390383>
- Hickman, C. (1997). The devil and the one drop rule: Racial categories, African Americans, and the U.S. census. *Michigan Law Review, 95*, 1161–1265. <http://dx.doi.org/10.2307/1290008>
- Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. *Memory & Cognition, 41*, 1021–1031. <http://dx.doi.org/10.3758/s13421-013-0316-7>

- Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology, 41*, 598–608. <http://dx.doi.org/10.1016/j.jesp.2004.10.007>
- Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*, 1168–1187. <http://dx.doi.org/10.1037/a0020463>
- Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). Overview of race and Hispanic origin: 2010. *2010 Census briefs*. Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>
- Hurlbert, A. (2001). Trading faces. *Nature Neuroscience, 4*, 3–5. <http://dx.doi.org/10.1038/82877>
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102*, 116–131. <http://dx.doi.org/10.1037/a0025335>
- Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review, 14*, 288–294. <http://dx.doi.org/10.3758/BF03194066>
- Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science, 11*, 379–385. <http://dx.doi.org/10.1111/1467-9280.00274>
- Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*, 89–94. <http://dx.doi.org/10.1038/82947>
- Levy, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open, 6*, 1–17. <http://dx.doi.org/10.1177/2158244016636433>
- Lew, T. F., & Vul, E. (2015). Structured priors in visual working memory revealed through iterated learning. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1332–1337). Austin, TX: Cognitive Science Society.
- Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings Biological Sciences, 272*, 2283–2287. <http://dx.doi.org/10.1098/rspb.2005.3220>
- Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008). Category contingent aftereffects for faces of different races, ages and species. *Cognition, 106*, 1537–1547. <http://dx.doi.org/10.1016/j.cognition.2007.06.008>
- Lyons, A., & Kashima, Y. (2001). The reproduction of culture: Communication processes tend to maintain cultural stereotypes. *Social Cognition, 19*, 372–394. <http://dx.doi.org/10.1521/soco.19.3.372.21470>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*, 1122–1135. <http://dx.doi.org/10.3758/s13428-014-0532-5>
- MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7*, 98–118. <http://dx.doi.org/10.1037/1076-8971.7.1.98>
- Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science, 24*, 2351–2360. <http://dx.doi.org/10.1177/0956797613495418>
- Marcus, G. F., & Davis, E. (2015). Still searching for principles: A response to Goodman et al. *Psychological Science, 26*, 542–544. <http://dx.doi.org/10.1177/0956797614568433>
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35. <http://dx.doi.org/10.1037/1076-8971.7.1.3>
- Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology, 97*, 405–423. <http://dx.doi.org/10.1348/000712605X85871>
- Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17*, 608–615. <http://dx.doi.org/10.1111/j.1467-9280.2006.01752.x>
- Ng, W. J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25*, 217–232. <http://dx.doi.org/10.1177/0022022194252004>
- Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., Jr., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96*, 795–810. <http://dx.doi.org/10.1037/a0013265>
- Peery, D., & Bodenhausen, G. V. (2008). Black + white = black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science, 19*, 973–977. <http://dx.doi.org/10.1111/j.1467-9280.2008.02185.x>
- Rhodes, G., & Jaquet, E. (2011). Aftereffects and reveal that adaptive face-coding mechanisms are selective for race and sex. In R. B. Adams, N. Ambady, K. Nakayama, & S. Shimojo (Eds.), *The science of social vision* (pp. 347–362). New York, NY: Oxford University Press.
- Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46*, 2977–2987. <http://dx.doi.org/10.1016/j.visres.2006.03.002>
- Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception, 38*, 232–241. <http://dx.doi.org/10.1068/p6110>
- Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*, 440–444.
- Sirovich, L., & Meytlis, M. (2009). Symmetry, probability, and recognition in face space. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 6895–6899. <http://dx.doi.org/10.1073/pnas.0812680106>
- Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10*, 479–491.
- Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS ONE, 7*, e41193. <http://dx.doi.org/10.1371/journal.pone.0041193>
- Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & Psychophysics, 69*, 619–627. <http://dx.doi.org/10.3758/BF03193919>
- Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*, B1–B9. <http://dx.doi.org/10.1016/j.cognition.2003.09.011>
- Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*, 42–50. <http://dx.doi.org/10.1109/38.946630>
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623–1626. <http://dx.doi.org/10.1126/science.1110589>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519–545. <http://dx.doi.org/10.1146/annurev-psych-113011-143831>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Ex-*

- perimental Psychology A: Human Experimental Psychology*, 43, 161–204. <http://dx.doi.org/10.1080/14640749108400966>
- Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 69, 1996–2019. <http://dx.doi.org/10.1080/17470218.2014.990392>
- Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, 20, 461–475. <http://dx.doi.org/10.1002/acp.1191>
- Walker, P. M., & Hewstone, M. (2008). The influence of social factors and implicit racial bias on a generalized own-race effect. *Applied Cognitive Psychology*, 22, 441–453. <http://dx.doi.org/10.1002/acp.1382>
- Wang, Y., & Zhou, L. (2016). Long-term exposure to American and European movies and television series facilitates Caucasian face perception in young Chinese watchers. *Perception*, 45, 1151–1165. <http://dx.doi.org/10.1177/0301006616652052>
- Waugh, L. (1982). Marked and unmarked: A choice between unequals in semiotic structure. *Semiotica*, 38, 299–318. <http://dx.doi.org/10.1515/semi.1982.38.3-4.299>
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428, 557–561. <http://dx.doi.org/10.1038/nature02420>
- Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 366, 1702–1725. <http://dx.doi.org/10.1098/rstb.2010.0360>
- Willenbockel, V., Fiset, D., & Tanaka, J. W. (2011). Relative influences of lightness and facial morphology on perceived race. *Perception*, 40, 621–624. <http://dx.doi.org/10.1068/p6963>
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, 42, 671–684. <http://dx.doi.org/10.3758/BRM.42.3.671>
- Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology*, 60, 107–126. <http://dx.doi.org/10.1016/j.cogpsych.2009.09.002>
- Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, 76, 1–29. <http://dx.doi.org/10.1016/j.cogpsych.2014.11.001>
- Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2009). Interracial contexts debilitate same-race face recognition. *Journal of Experimental Social Psychology*, 45, 1123–1126. <http://dx.doi.org/10.1016/j.jesp.2009.05.009>
- Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the Cross-Race Effect. *Personality and Social Psychology Review*, 16, 116–142. <http://dx.doi.org/10.1177/1088868311418987>
- Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, 8, 161–185. <http://dx.doi.org/10.1521/soco.1990.8.2.161>

Received June 14, 2017

Revision received January 31, 2018

Accepted March 20, 2018 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!