Normalizing Anomalies with Mobile Exposure (NAME): Reducing implicit biases against people with facial anomalies

Nadir Bilici[1-3]*, Clifford I. Workman[1-3]*†, Stacey Humphries[1-3],
Mariola Paruzel-Czachura[1-4], Roy H. Hamilton[1,2,5], & Anjan Chatterjee[1-3]

[1] Department of Neurology, University of Pennsylvania, Philadelphia, PA
[2] Penn Brain Science Center, University of Pennsylvania, Philadelphia, PA
[3] Penn Center for Neuroaesthetics, University of Pennsylvania, Philadelphia, PA
[4] Institute of Psychology, University of Silesia in Katowice
[5] Laboratory for Cognition and Neural Stimulation, University of Pennsylvania, Philadelphia, PA
* Joint first authorship

† Correspondence:
University of Pennsylvania
3710 Hamilton Walk
Goddard Laboratories, Room 332
Philadelphia, PA 19104
E-mail: cliffworkmanphd@gmail.com

Running Head: NORMALIZING ANOMALIES WITH MOBILE EXPOSURE

Keywords: implicit bias; disfigurement; faces; discrimination; stigma

Abstract: 118 words **Main Text:** 6,591 words **References:** 60 **Figures & Tables:** 5 & 2

Abstract: This pre-registered study tested whether implicit biases towards people with anomalous facial features, like scars and palsies, are reducible through routine exposure to anomalous faces bearing such anomalies. Implicit biases were measured before and after completing an exposure intervention—either to facial anomalies, or to people of color (POC). The interventions were delivered remotely via mobile phone application and consisted of two sessions daily over 5 consecutive days. Sessions began with exposure to 10 different faces (anomalous or POC) followed by another face (anomalous or POC) presented alongside a moral exemplar story about them. The anomalous faces intervention reduced implicit biases against anomalous faces but not against POC. The POC intervention did not have targeted effects on implicit biases.

Keywords: implicit bias; disfigurement; faces; discrimination; stigma

Introduction

Most people encounter individuals with facial differences infrequently in daily life (Workman et al., 2021). This fact is not surprising, given a prevalence of facial differences in approximately 1 out of every 100 people (Julian & Partridge, 2008). The available evidence suggests that people go out of their way to *avoid* interacting with people who look different (Bull & Stevens, 1981; Dey et al., 2015; Houston & Bull, 1994; Rumsey et al., 1982). Recent research indicates that an "anomalous-is-bad" stereotype is expressed in biases against people with facial anomalies (Hartung et al., 2019; Jamrozik et al., 2019; Workman et al., 2021). Given the limited opportunities people have to form impressions about individuals with facial anomalies—negative or otherwise—the origins of biases against them are unclear. Hollywood villains are more likely than heroes to have facial anomalies (Croley et al., 2017), like the *Lion King's* Scar. Since the representation of facial anomalies in popular culture takes for granted that flaws to one's face imply flaws to one's character, this received wisdom, perhaps unsurprisingly, permeates attitudes about facial anomalies. There is an urgent need for evidence-based interventions to attenuate negative biases against people with facial anomalies, especially since their being stigmatized can cause psychological trauma that may lead to suicide (Biernoff, 2011).

Why should a person's appearance evoke inferences about their character? These inferences may be a nasty side effect of an adaptive aversion to the threat of pathogen exposure. Although avoiding diseased others can protect us from infection, sensitivity to pathogen exposure may overgeneralize to features that do not themselves signal poor fitness (Zebrowitz & Montepare, 2008). Consistent with this view, facial anomalies can arouse feelings of disgust in highly sensitive observers that scale with the severity of the anomalies themselves—the more severe the anomaly, the stronger the disgust response (Shanmugarajah et al., 2012). People also

show similar disgust reactions when faced with the prospect of handling objects they believed were last touched by people with flu symptoms and by people with noninfectious facial anomalies (Ryan et al., 2012). In other words, facial anomalies can be treated like expressions of infectious disease. A recent cross-cultural study with the Hadza hunter-gatherers of Tanzania casts doubt on an overgeneralized pathogen avoidance account, however, suggesting instead that the anomalous-is-bad stereotype is learned (Workman et al., 2022). Repeated exposure to negative depictions of people with anomalous faces—in movies and other media, for instance—may strengthen reflexive negative emotional responses elicited by anomalous faces that drive negative affective behaviors like social avoidance (Amodio & Cikara, 2021). This prospect raises an important question: Can repeated exposure to *positive* representations of people with facial anomalies as moral heroes undermine biases against them?

Stories about moral heroes are widely used in the context of moral education to inspire changes in emotional reactions and behaviors (Bandura, 1969; Kristjánsson, 2006; Sanderse, 2013). This research has focused on characterizing the features of our psychology that maximize the emulation of moral exemplars. Less is known about how moral exemplar stories shape our impressions of the exemplars themselves. Nevertheless, recent evidence suggests our impressions about exemplars—like their similarity to us and whether we are capable of their actions—are critical to the efficacy of those stories in inspiring morally good behavior (Han et al., 2017, 2022). Even morally irrelevant information may attenuate negative biases. Participants faced with harming another person in a hypothetical moral scenario, for instance, were less willing to do so after taking that person's perspective in a non-moral context (e.g., deciding what to order at a restaurant; Majdandžić et al., 2012). In fact, even very rapid and brief exposures of white participants to faces of racial out-group members—without any accompanying

information, moral or otherwise—increased the liking of novel faces from those same groups (Zebrowitz et al., 2008).

We present the "Normalizing Anomalies with Mobile Exposure" (NAME) intervention, which combines moral exemplar stories with exposure to people with anomalous faces in order to mitigate biases against them. Although there is some evidence for both implicit *and* explicit biases against people with facial differences (Hartung et al., 2019; Workman et al., 2021), this intervention was specifically designed to undermine negative implicit attitudes. Implicit bias recognition and management curricula are increasingly used to advance equity (Sukhera et al., 2020) and are effective in improving self-awareness of implicit biases (Zeidan et al., 2019), in changing negative implicit attitudes (Leslie et al., 2018), and in reducing discomfort with disability (Hearn & Hearn, 2020). In this study, participants' implicit biases were measured before and after they completed one of two possible exposure interventions—to people with facial anomalies, or to people of color (POC). The intervention, which was delivered remotely using a custom mobile phone application, was comprised of two sessions per day over 5 consecutive days. Each session began with exposure to 10 different face photographs (anomalous or POC). Then, another person's face (anomalous or POC) was presented alongside a moral exemplar story ostensibly about them. Before each session ended, participants answered questions about the story and how it made them feel.

This pre-registered study (https://osf.io/b9g6v/) tested the hypothesis that a lack of exposure to people with facial anomalies underpins implicit biases against them. Several predictions follow from this hypothesis. First, repeated exposure to positive representations of people with facial anomalies or to POC should weaken implicit biases against them. Second, the exposure intervention is expected to be more effective at weakening implicit biases against facial

anomalies than POC. Although both groups are underrepresented and marginalized in American popular culture (Erigha, 2018), exposure to facial anomalies is nevertheless expected to be more effective. The odds are stacked against encountering people with facial anomalies—as a consequence, the impressions we form about them often derive from little more than superficial information. In contrast, encounters with POC are much more likely to occur in daily life, especially in metropolitan areas like those where the research was conducted, and positive representations of POC as moral heroes are increasingly common (White, 2018). White participants in one study reported feeling more negatively than POC about past instances of interracial contact, negatively impacting the quality and quantity of subsequent instances of interracial contact (Doerr et al., 2011). Biases against POC may be more deeply ingrained—that is, less susceptible to the mechanism exploited by the intervention—than the kinds of biases detected against people with facial anomalies.

Method

Participants

A sample of $N = 100$ healthy adult volunteers was recruited to participate in this study (see Table 1 for demographics). Participants were monetarily compensated for their time—compensation was dispensed after participants finished the second of two study visits. Eligibility was contingent on having access to a smartphone (iPhone or Android) with sufficient disk space to load our phone application and regular internet access to allow the app to make secure backups of de-identified participant data. Participants were also required to speak English fluently and could not have significant visual impairments that might interfere with the study procedures. No participants were excluded from further analysis. This study was conducted in a manner

consistent with the American Psychological Association's Ethical Principles in the Conduct of Research with Human Participants. The study procedures were approved by the Institutional Review Board at the University of Pennsylvania (Protocol #806447). Participants gave written informed consent before starting any of the study procedures and were prompted to re-confirm their consent on first opening the phone application.

Participants enrolled in the study through an online participant pool that is heavily but not exclusively comprised of undergraduate students. Participants were required to schedule baseline and follow up study visits occurring within 5-8 days of one another (1 participant returned after 9 days and another after 10, but all other follow-up visits took place within this 8-day window). Between visits, participants completed the smartphone-based "Normalizing Anomalies with Mobile Exposure" (NAME) intervention. Each session of the intervention had two parts. In the first part, adapting methods from Zebrowitz and colleagues (2008), we leveraged the mere-exposure effect by exposing participants to a succession of faces from a stigmatized social group. In the second part, based on work by Majdandžić and colleagues (2012), participants read a short story about a moral exemplar—ostensibly from the same stigmatized social group—and were encouraged to represent their mental states.

An *a priori* power analysis was conducted in G*Power ($\alpha = 0.05$, $1\text{-}\beta = 0.8$; Faul et al., 2007) using effect sizes from the two studies adapted for the intervention (Majdandžić et al., 2012; Zebrowitz et al., 2008). Zebrowitz and colleagues (2008) reported two relevant medium-sized effects ($d_z = 0.45, 0.54$), whereas the effect detected by Majdandžić and colleagues (2012) was smaller in magnitude ($d_z = 0.27$). The power analyses based on the effects from Zebrowitz and colleagues (2008) indicated $N = 42$ participants per intervention would provide adequate statistical power. Using the effect reported by Majdandžić and colleagues (2012), however,

Table 1 | The demographics characterizing the entire sample of participants enrolled in the NAME Study and characterizing each of the two interventions separately.

	NAME Study Participants (*N* = 100)				**Anomalous Faces Intervention** (*N* = 50)				**POC Faces Intervention** (*N* = 50)			
	M	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
Age	24.33	7.71	18	56	23.36	6.28	18	44	25.30	8.88	18	56
Education (Years)	15.09	2.17	12	24	14.82	1.70	12	18	15.36	2.55	12	24
Political Views – Social	5.49	1.29	2	7	5.56	1.36	2	7	5.42	1.23	2	7
Political Views – Economic	4.85	1.50	1	7	4.78	1.62	1	7	4.92	1.38	2	7
IRI - Perspective Taking	3.79	0.54	2.43	5.00	3.77	0.51	2.71	5.00	3.82	0.57	2.43	5.00
Average Number of Days Between Study Visits	6.74	1.05	4	10	6.74	1.17	4	10	6.74	0.92	5	8
Race / Ethnicity (*N*)												
White	46				24				22			
Asian	31				15				16			
Black	13				6				7			
Hispanic or Latinx	3				3				0			
Multiracial	5				2				3			
Other / do not wish to say	2				0				2			
Biological Sex (*N*)												
Male	25				14				11			
Female	75				36				39			
Other / do not wish to say	0				0				0			
Gender (*N*)												
Male	24				14				10			
Female	74				35				39			
Other / do not wish to say	2				1				1			
Sexuality (*N*)												
Heterosexual	79				40				39			
Homosexual	4				1				3			
Bisexual	9				5				8			
Other / do not wish to say	8				4				0			

The Interpersonal Reactivity Index (IRI) measured dispositional perspective taking, alongside several other facets of empathy (Davis, 1983), on a 5-point scale from "does not describe me well" to "describes me very well." Social and economic political views were assessed with a 7-point scale ranging from less to more conservative—these values were then reverse scored, such that larger scores indicate more liberalism (and smaller scores more conservatism). Econ, Economic; IRI, Interpersonal Reactivity Index; NAME, Normalizing Anomalies with Mobile Exposure.

indicated sample sizes closer to $N = 100$ participants per intervention could be necessary. For efficient sampling, an optional stopping rule was used: if a sample of $N = 50$ per intervention approached but did not reach significance ($p = 0.05$ - $p = 0.1$), a second wave of recruitment would double the samples for each intervention, and a more stringent Bonferroni corrected $\alpha = (0.05 / 2) = 0.025$ would be employed thereafter. A second wave of recruitment was not necessary, with only $N = 50$ participants ultimately required for each intervention. The sample size justification for this study, including the power analyses and optional stopping rule, was pre-registered: https://osf.io/b9g6v/.

Study procedures

Materials, de-identified data, code, and statistical outputs are publicly available from: https://osf.io/vm7d4/.

Psychological dispositions

Before their baseline study visit, participants completed a brief online survey through Qualtrics, which included self-report measures—described in greater detail in the Supplementary Online Materials—as well as demographic questions. Four self-report measures quantified individual differences in beliefs about justice and egalitarianism and in sensitivity to disgust and to different facets of empathy (e.g., cognitive empathy or perspective taking). Since previous work points specifically to perspective taking as an effective way to improve morally-relevant attitudes about

other people (Majdandžić et al., 2012), a subscale from the Interpersonal Reactivity Index (IRI) that indexes dispositional perspective taking is examined below (Davis, 1983). The other self-report measures were acquired for exploratory purposes and are not explored further here.

Implicit and explicit biases

The primary outcome variable was change in implicit bias after the NAME intervention. Implicit bias is the pre-reflective attribution of characteristics to other people based on their group memberships (Jost et al., 2009) and was assessed with the Implicit Association Test (IAT; Greenwald et al., 1998). Over 20 years after its initial publication, the IAT remains a widely used instrument used to infer the strength of unconscious intergroup attitudes (Banaji & Greenwald, 2016; Kurdi et al., 2019). Recent research with the IAT finds that people harbor robust implicit biases against individuals with visible facial anomalies (Changing Faces, 2017; Hartung et al., 2019). Although not without controversy (e.g., Oswald et al., 2013), we used the IAT because it has the most favorable psychometric properties among several alternative assessments (Bar-Anan & Nosek, 2014).

At the baseline study visit, participants completed two versions of the IAT. In both versions, they associated photographs of faces with "good" and "bad" words (good: adore, approachable, attractive, excellent, friendly, happy, lovely, and spectacular; bad: despise, disaster, disgust, evil, pain, rotten, sickening, and ugly). In the first version, the photographs were of faces that either did or did not have visible anomalies. In the second version, the faces—none with visible anomalies—were either white or belonged to POC. Participants repeated both IATs at the follow-up study visit. Larger IAT *d*' scores—detected in people who more rapidly linked anomalous (or POC) faces to bad words than good— signify stronger implicit bias (the

scoring algorithm used to compute the IAT *d*' is given in the Supplementary Materials). Participants also completed the Explicit Bias Questionnaire (EBQ) before and after the intervention to detect changes to explicit biases against people with facial anomalies (Hartung et al., 2019; Workman et al., 2021). The EBQ is a 33-item scale that assesses personal history with and overall beliefs about people with facial anomalies, with larger EBQ scores denoting stronger explicit bias.

Normalizing Anomalies with Mobile Exposure (NAME) intervention

After completing the baseline IATs and EBQ, participants were pseudo-randomly assigned to one of two versions of the NAME intervention. In the "anomalous faces" version ($N = 50$), participants were shown photographs of people with visible facial anomalies. In the "POC" version ($N = 50$), participants saw photographs of faces belonging to POC. The demographic makeup of the samples assigned to each intervention is given in Table 1. Participants were prompted to complete two intervention sessions each day (once in the morning and once in the afternoon, each lasting around 3 minutes) over 5 contiguous days for a maximum of 10 sessions (around 30 minutes). Participants received daily push notifications on their mobile devices at 8am and again at 12:30pm to remind them to complete these sessions. The completion rate for all sessions across participants was 91.4% for an average of less than one missed session per participant. No participants were excluded for missing over 50% of the intervention sessions (6 or more missed sessions). All participants returned for their follow-up visit within 1.75 ± 1.03 days of completing the intervention.

Smartphone application

The smartphone application was developed by N.B. with React Native (https://reactnative.dev/) in Expo (https://expo.io/) for the iOS and Android mobile operating systems. The source code (JavaScript) can be obtained from: https://osf.io/vm7d4/. The baseline study visit concluded after participants successfully installed the NAME application onto their mobile devices and completed a practice session of the intervention. At the end of this and all subsequent intervention sessions, the application transmitted newly recorded participant responses to a secure server (REDCap; https://www.project-redcap.org/). The graphical user interface for the NAME smartphone application is illustrated in Figure 1.

Face photographs

In the first part of each of the 10 intervention sessions, participants were exposed to 10 non-repeating face images (anomalous or POC depending on intervention; Figure 1C). Each intervention therefore required a total of 100 face images (10 non-repeating face images * 10 sessions = 100 non-repeating face images in total). On seeing each face, participants judged the biological sex of the face to be either male or female. Between each of the 10 faces per session, participants saw a black screen for 500ms.

For the anomalous faces version of the NAME intervention, 100 images of anomalous faces (56 female; 7 Asian, 3 Black, 7 Latinx, and 82 white; Table S2) were selected from the Chatlab Facial Anomaly Database (CFAD; Workman & Chatterjee, 2021). For the POC version of the intervention, 100 face photographs (47 female; 28 Asian, 38 Black, 34, Latinx, 0 white; no visible facial anomalies; Table S2) were selected from the Chicago Face Database (Ma et al., 2015). Additional information about the face photographs is provided in the Supplementary Materials.

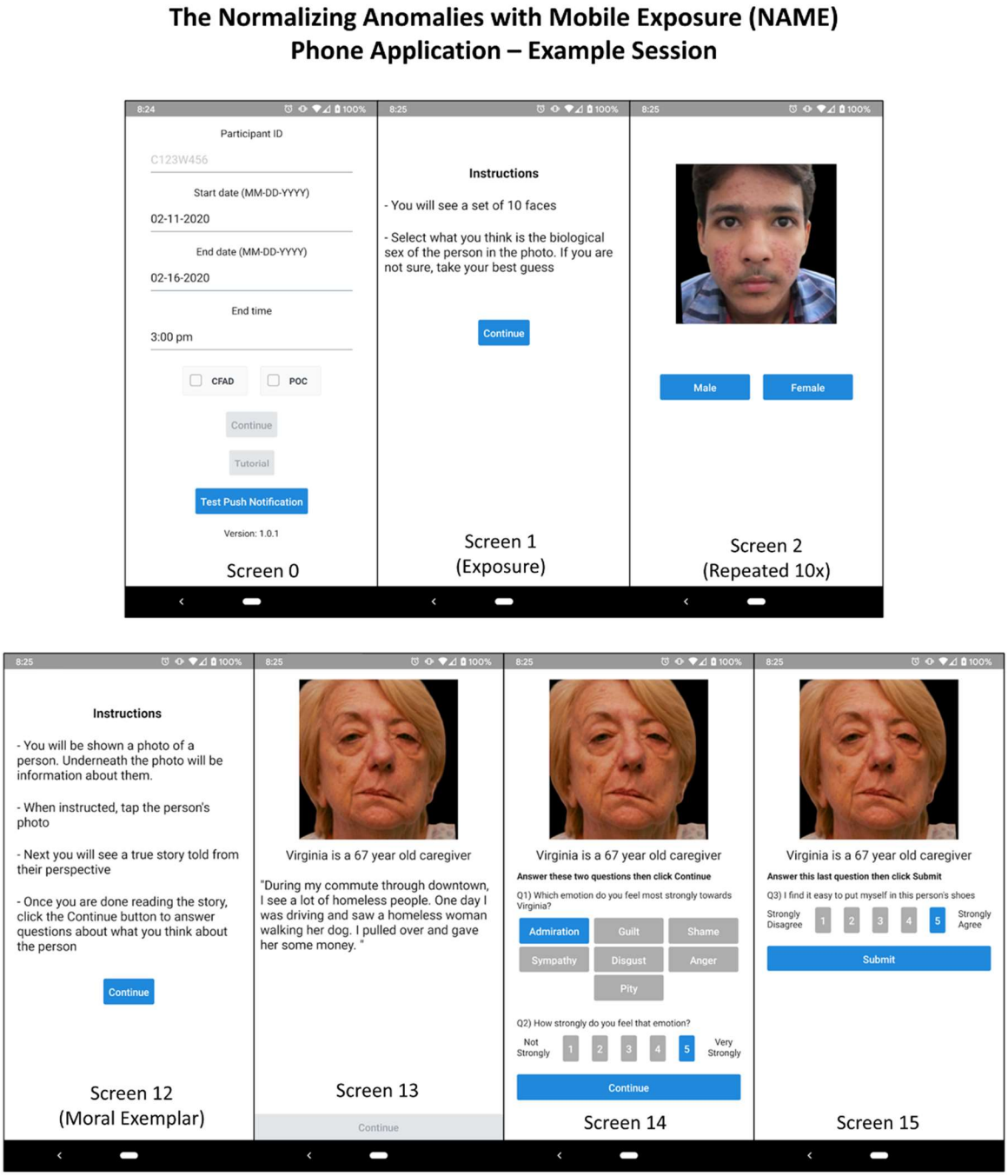

Figure 1 | An example session from the "Normalizing Anomalies with Mobile Exposure" (NAME) phone application. After being signing into the NAME phone application by an experimenter (Screen 0), participants completed the first of 10 sessions of the NAME intervention. Each session, delivered twice daily over the course of five days, was comprised of two parts. In the first part (Screens 1 through 11), participants were exposed to a succession of faces from a stigmatized social group (either anomalous faces or people of color). In the second part (Screens 12 and 13), participants read a short story about a moral exemplar—ostensibly from the same stigmatized social group—and were encouraged to represent the exemplar's mental states (Screens 14 and 15).

Moral exemplar characteristics

In the second part of each session, participants learned four pieces of information ostensibly about a moral exemplar before reading a story about the exemplar's praiseworthy actions (Figure 1E). The first piece of information was the moral exemplars' face—a single photograph was chosen at random to represent the moral exemplar. For the anomalous faces intervention, this photo was one of 50 from the 100 CFAD images with complete or near-complete demographic information (non-repeating; Table S2). For the POC intervention, all 100 images were used in both parts of each session.

After seeing the moral exemplar's photo, participants were given the remaining three pieces of information: the exemplar's name, age, and occupation (Figure 1E-1G). A pool of 200 names comprised of 100 stereotypically male names and 100 stereotypically female names was used in this study (Table S1). The male and female names were randomly assigned to the fictional male and female moral exemplars that were shown to participants as part of the NAME intervention. According to the United States Social Security Administration, these represent the most popular baby names from 1919 to 2018[1]. Common names enhanced the relatability of moral exemplars. Recent evidence finds the power of moral exemplars to inspire others is contingent on how relatable people find the exemplars to be (Han et al., 2017, 2022).

The age of each exemplar—provided in the documentation for the Chatlab Facial Anomaly and Chicago Face Databases—was presented alongside the exemplar's name and supposed occupation. We selected the 5 occupations from Jenkins and colleagues (2018) that elicited the most favorable ratings along dimensions of warmth (what do others intend?) and competence (are they capable of acting on their intentions?): athlete, farmer, nurse, parent, and

[1] Permalink: https://web.archive.org/web/20200820060122/https://www.ssa.gov/oact/babynames/decades/century.html

surgeon. To limit repetition, we also devised and included six variations of each of occupation—in addition to “athlete”, for instance, the list included “pro-gamer”, “sports coach”, and “gymnast” (see Table S3 for the complete list).

Moral exemplar stories

Next, the photograph and characteristics of the moral exemplar were paired with a vignette ostensibly describing the person’s praiseworthy actions. Forty-one vignettes recounting exemplary moral actions from the first-person perspective were identified from a larger set of standardized moral scenarios about the real-world experiences of actual people (Knutson et al., 2010). Five example moral exemplar stories used in the NAME intervention are shown in Table 2. After each vignette, participants answered questions about the moral exemplar and their feelings towards them. Specifically, participants answered the following 3 questions (Figure 1F & 1G): “Which emotion do you feel most strongly towards [exemplar name]?” (admiration, disgust, guilt, shame, sympathy, anger, or pity), “How strongly do you feel that emotion?” (1 = *Not strongly*, 5 = *very strongly*), and “I find it easy to put myself in this person’s shoes” (1 = *Strongly disagree*, 5 = *Strongly agree*). As in Majdandžić et al. (2012), merely answering these questions was expected to attenuate implicit bias by priming participants to take the perspectives of moral exemplars from stigmatized social groups.

Each intervention session concluded with a screen that reminded participants about their next session, and their follow-up in person visit. At this time, participant data was automatically transmitted to a secure REDCap database. Each session took approximately 3 minutes to complete, such that the entire intervention lasted around 30 minutes.

Table 2 | Examples of 5 moral exemplar stories that were used as part of the NAME intervention.

	Story	Self-Benefit	Other-Benefit	Emotional Intensity
1	I was running across the street one time, years ago, and a little old woman fell. She was carrying groceries and she slipped. I helped her get her bearings back and picked up the groceries.	3.57	6.13	3.57
2	I was friends with a woman who wanted to pursue her acting career. She was going to pawn this ring that was in her family for years. I didn't want her to pawn the ring so I gave her $200 and said that I would hold on to it for her until she could pay me back.	2.80	5.40	3.87
3	I bought a pair of shoes and the lady was busy, so she charged me the wrong price for the shoes. I didn't want to cheat her so asked her if she was sure that this was the right price for the shoes. It was the wrong price and she corrected the price and thanked me.	3.17	5.53	2.83
4	I was at the pharmacy buying something and I noticed a man who was sitting outside selling trinkets. He was homeless and it was freezing out. So I went next door to a store and bought him some food and new clothes.	3.33	6.30	4.50
5	I knew a couple and their child through one of those "save-a-child" foundations. The family decided to move to California and I knew they were going to have a really hard time finding work. So I sold a bracelet that I had and sent them the money.	3.30	6.17	4.70

All of the stories were drawn from an earlier study (Knutson et al., 2010). As part of this earlier study, participants rated the vignettes along a number of dimensions, including self-benefit, other-benefit, and emotional intensity (1 = "Not at all", 7 = "Extremely"). The prosociality captured by the vignettes is evident in the larger scores each of these vignettes received for other- relative self-benefit. Since we were interested in how variability in affective responses to the moral exemplar stories might affect the intervention's efficacy, it was important that stories evoke a range of emotional reactions. A column for emotional intensity is also given and underscores the variability in emotional responses elicited by the vignettes. NAME, Normalizing Anomalies with Mobile Exposure.

Statistical analyses

Sample size and significance

The sample size was determined with a power analysis based on an effect reported in a previous study using mere exposure to out-group faces to reduce implicit biases ($d_z = 0.44$, $\alpha = 0.05$, and $1\text{-}\beta = 0.8$; Zebrowitz et al., 2008). This power analysis indicated that sample sizes of $N = 42$ per group would provide sufficient power to detect the effect of interest with oversampling ($N = 50$ per group) to account for attritions or exclusions of poor-quality data.

Another relevant study used a perspective taking manipulation similar to the one employed in this study to demonstrate a smaller but nevertheless significant effect (d_z = 0.27 Majdandžić et al., 2012). A second power analysis based on this effect size suggested samples of N = 102 per group would provide sufficient power. In the event that samples of N = 50 per group yielded results nearing but not obtaining significance (p = 0.05 - p = 0.1), a second wave of recruitment was planned to increase the sample sizes per group in line with the second power analysis. To account for the interim analysis, a Bonferroni corrected α = (0.05/2) = 0.025 was planned for determining significance. The optional stopping procedure was pre-registered (https://osf.io/b9g6v/). Ultimately, samples of N = 50 were sufficient and a second wave of recruitment was not carried out.

Statistical models

The analysis plan was also pre-registered (https://osf.io/b9g6v/). All analyses were conducted in JASP (JASP Team, 2019). The primary outcome variable was change in implicit biases before versus after the intervention. A repeated measures t-test examined whether implicit biases congruent with the intervention type (e.g., IAT scores reflecting implicit biases against people with anomalous faces in those participants who underwent the anomalous faces intervention) were significantly reduced following the intervention. This test indicated whether the intervention paradigm—regardless of intervention type—reduced the corresponding IAT measures of implicit biases as expected.

Next, to test whether the efficacy of the intervention differed as a function of intervention type, a two-by-two mixed ANOVA was conducted with intervention type as a between-subjects variable (anomalous faces or POC) and intervention congruent IAT scores (pre and post) as a

repeated-measures variable. This test indicated whether the interventions independently reduced corresponding IAT measures of implicit bias as expected.

Next, to examine the specificity of the interventions, a two-by-two repeated-measures ANOVA was conducted that accounted for scores from all four IATs independently—both pre and post, and both POC and anomalous faces. Targetedness was measured as the effect of change in intervention congruent IAT *d*' scores compared to incongruent IAT *d*' scores. The predicted interaction was that scores from the congruent IAT should show greater change than the scores from the incongruent IAT.

Finally, we examined whether prosocial affect (indexed with questions about emotional responses to the intervention) or cognitive empathy (indexed with a question about ease of perspective taking) in response to the exposure intervention best predicted the efficacy of the intervention in reducing congruent implicit biases. First, we calculated the frequency with which participants experienced prosocial emotions during the intervention (sympathy, admiration, guilt, pity). Then, we examined whether the difference score reflecting changes in congruent biases correlated with the average strength of prosocial emotions experienced throughout the intervention. This analysis was repeated for the antisocial moral emotions (disgust, anger, and shame). An additional correlation examined relations between average cognitive empathy and changes in bias. The resulting correlation coefficients were *z*-transformed and were then compared to determine whether one correlation was significantly strongly than the other.

Results

Pre-registered analysis 1: Did the overall intervention reduce targeted implicit biases?

To examine the overall effect of the NAME intervention, the data from the anomalous face and POC versions of the intervention were pooled. Implicit biases that were congruent with the type of intervention the participant received were significantly reduced at time 2 (IAT *d*' = -0.22 ± 0.57) relative to time 1 (*d*' = 0.03 ± 0.48; $t(97) = 3.96$, $p < .001$, 95% CI [0.13, 0.38], $d = 0.40$; see Figure 2).

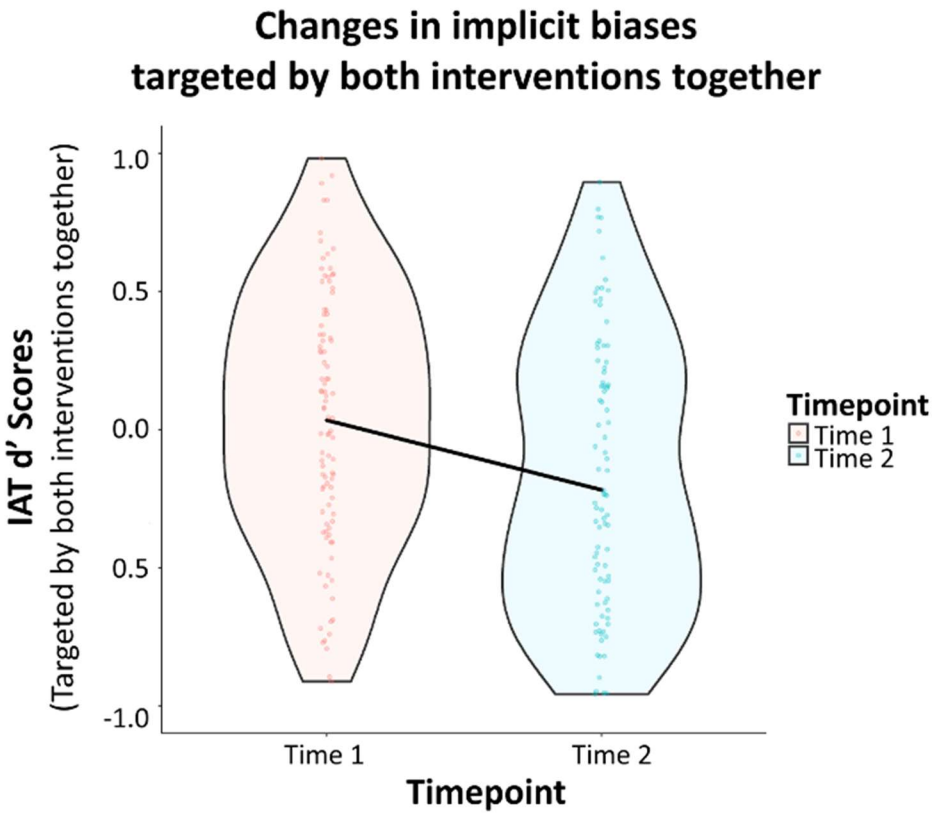

Figure 2 | Changes in congruent implicit biases targeted by both interventions collapsed together. Implicit biases that were congruent with the type of intervention the participant received (i.e., implicit biases against anomalous faces in participants who underwent the anomalous faces intervention and implicit biases against people of color in participants who underwent the people of color intervention) were significantly reduced at time 2 relative to time 1 regardless of which specific intervention participants completed. This suggests that the NAME intervention was effective in reducing congruent implicit biases.

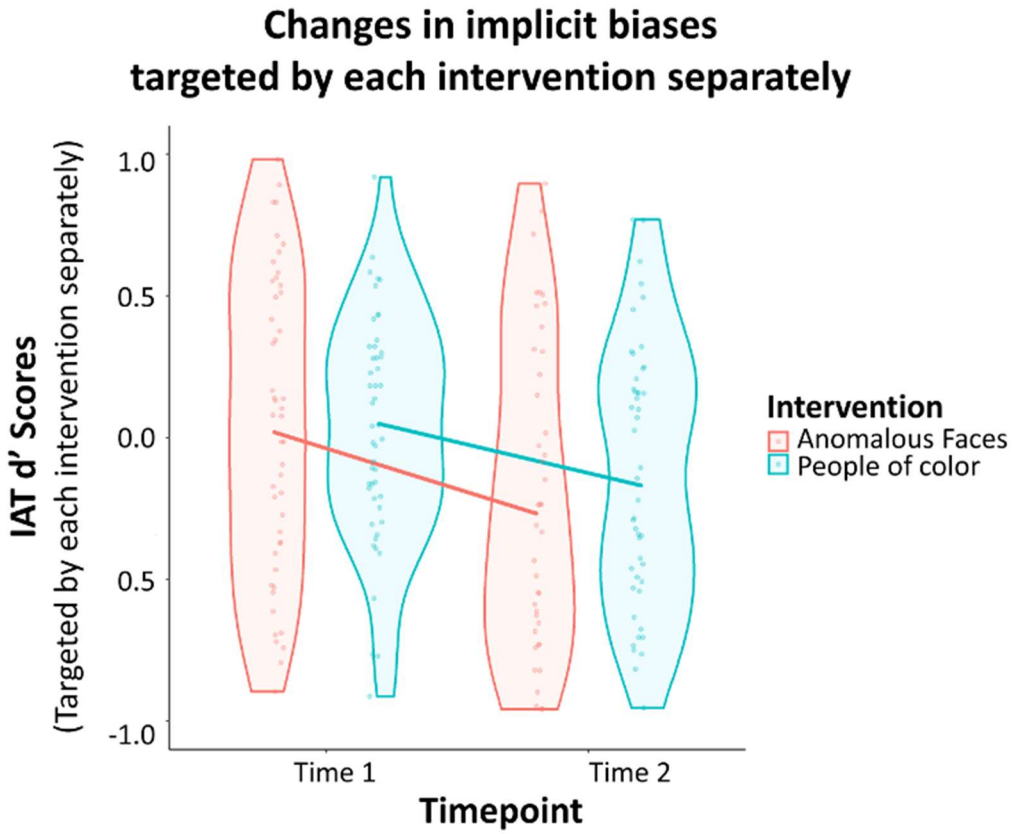

Figure 3 | Changes in congruent implicit biases targeted by each intervention separately. Implicit biases that were congruent with the type of intervention the participant received were significantly reduced at time 2 relative to time 1 for both interventions separately. This suggests that the anomalous faces and people of color variants of the NAME intervention were both effective in reducing congruent implicit biases.

Pre-registered analysis 2: Did either intervention produce more robust effects than the other?

Next, we examined whether the two intervention types differed in their efficacy at reducing congruent implicit biases. A 2x2 mixed ANOVA revealed a main effect of time (as above, such that biases were significantly reduced at time 2 [time 1: anomalous faces intervention d' = 0.02 ± 0.54, POC intervention d' = 0.05 ± 0.42; time 2: anomalous faces d' = -0.27 ± 0.64, POC d' = -0.17 ± 0.48): $F(1, 96) = 15.61$, $p < .001$, $\eta^2 = 0.06$]). There was no main effect of the type of intervention (POC vs. anomalous faces: $F(1, 96) = 0.65$, $p = .42$, $\eta^2 = 0.01$, and no interaction

between intervention and time: $F(1, 96) = 0.43$, $p = .51$, $\eta^2 = 0.00$, suggesting that there were no differences in how effective the interventions were at reducing congruent biases (see Figure 3).

Pre-registered analysis 3: Did either intervention have nonspecific effects on untargeted implicit biases?

To test whether either intervention had nonspecific effects generalizing to both congruent and incongruent biases, we examined the data from each intervention separately. For the anomalous faces intervention, there was a main effect of time:$F(1, 48) = 6.49$, $p = .01$, $\eta^2 = 0.03$, no main effect of bias type: $F(1, 48) = 0.169$, $p = .68$, $\eta^2 = 0.00$, and a significant interaction between time and bias type: $F(1, 48) = 4.71$, $p = .04$, $\eta^2 = 0.02$. After receiving the anomalous faces intervention, implicit biases against anomalous faces were significantly reduced (time 1 *d*': 0.02 ± 0.54, time 2 *d*': -0.27 ± 0.64; $t = 3.35$, $p_{holm} = .007$, $d = 0.48$) but implicit biases against POC were not affected (time 1 *d*': -0.08 ± 0.45, time 2 *d*': -0.13 ± 0.40; $p_{holm} > .28$, $d < 0.28$; see Figure 4). Interestingly, despite a reduction in implicit biases against anomalous faces, explicit biases were not significantly reduced by the intervention (time 1 EBQ scores: 3.91 ± 0.57, time 2 EBQ scores: 3.91 ± 0.56; $t[49] = -0.12$, $p = .78$, 95% CI [-0.25, ∞], $d = -0.02$).

For the POC intervention, there was a main effect of time ($F(1, 48) = 5.42$, $p = .02$, $\eta^2 = 0.03$), a main effect of bias type ($F(1, 48) = 4.62$, $p = .04$, $\eta^2 = 0.02$), and no significant interaction between time and bias type ($F(1, 48) = 0.17$, $p = .68$, $\eta^2 = 0.00$). In contrast to participants assigned to the anomalous face intervention, participants assigned to the POC intervention demonstrated significantly greater implicit biases against POC relative to anomalous faces both before (*d*' = 0.05 ± 0.42) and after the intervention (*d*' = -0.17 ± 0.49). In addition, the significant main effect of time and the lack of an interaction between time and bias type suggests

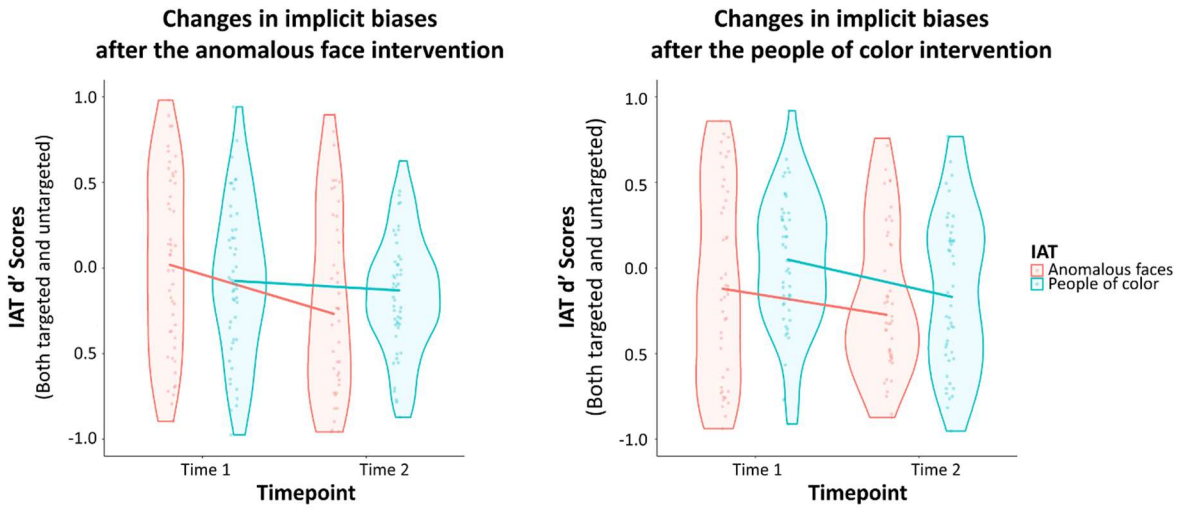

Figure 4 | Changes in both congruent and incongruent implicit biases targeted by each intervention separately. After receiving the anomalous faces intervention, congruent implicit biases against anomalous faces were significantly reduced but incongruent implicit biases against people of color were not affected. Participants assigned to the people of color intervention demonstrated significantly greater implicit biases against people relative to anomalous faces, and the people of color intervention reduced both types of bias and did not have a targeted effect on bias against people of color.

that the POC intervention reduced both types of bias (anomalous faces intervention time 1 *d*' = -0.12 ± 0.64, time 2 *d*' = -0.27 ± 0.50) and did not have a targeted effect on implicit biases against POC (see Figure 4).

Pre-registered analysis 4: Did prosocial affect or perspective taking relate to the efficacy of the intervention?

Finally, we examined how the participants' emotional responses and their self-reported ability to take the exemplar's perspective predicted the efficacy of the intervention. While participants were able to select from a list of both prosocial (e.g., admiration) and antisocial emotions (e.g., disgust), they overwhelmingly selected prosocial emotions. However, the strength of prosocial affect reported did not significantly relate to reductions in implicit bias (r = .11, p = .14, 95% CI [-.06, 1.00]). Perspective taking ratings given during the intervention did significantly correlate

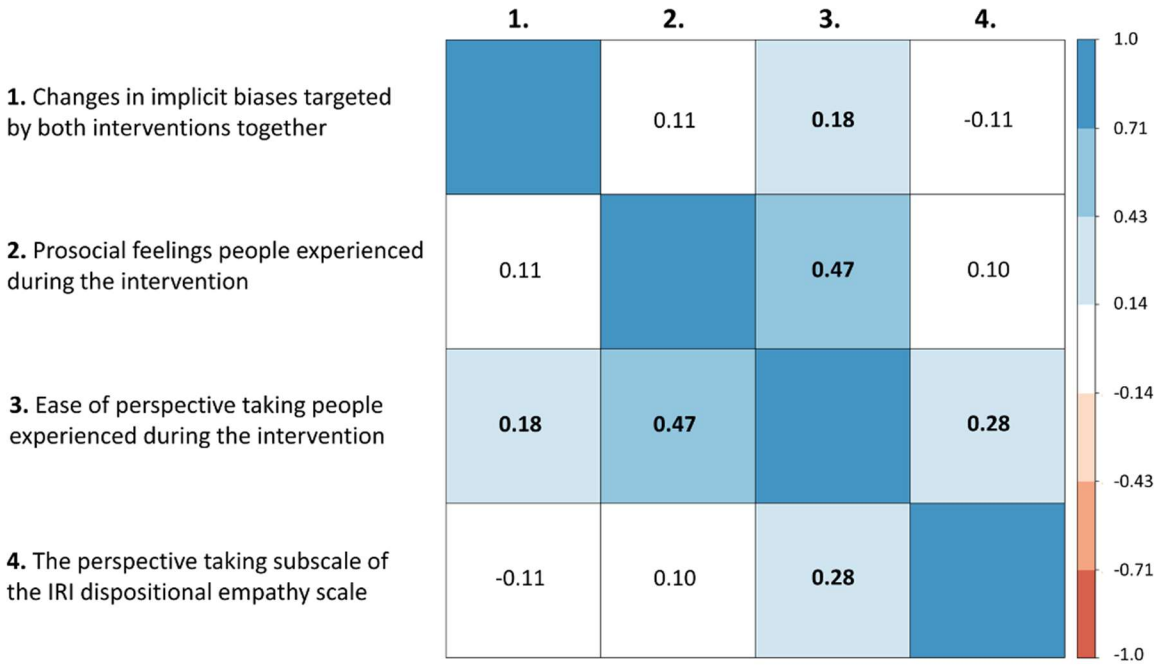

Figure 5 | Correlations between changes in implicit biases, prosocial affect, and trait and state perspective taking. Reductions in implicit biases were positively correlated with perspective taking ratings made during the intervention, but were not associated with prosocial affect ratings made during the intervention nor with trait perspective taking measured with the Interpersonal Reactivity Index (Davis, 1983).

with reductions in bias (r = .18, p = .03, 95% CI [.02, 1.00]), but trait levels of perspective taking measured by the IRI did not (r = -.11, p = .85, 95% CI [-.27, 1.00], see Figure 5).

Discussion

Can exposure to people with facial anomalies, like scars or port wine stains, reduce implicit biases towards them? This pre-registered study (https://osf.io/b9g6v/) tested the hypothesis that implicit biases towards people with facial differences occur, in part, because they are encountered infrequently by people with typical faces. If true, implicit bias should be reduced through routine exposure to faces bearing anomalous features. A novel intervention—"Normalizing Anomalies with Mobile Exposure" (NAME)—was designed and implemented in the present study. Before completing the NAME intervention, participants' implicit biases

against people with visible facial differences and against people of color (POC) were measured. Participants were then randomly assigned to one of two exposure interventions—either to people with facial anomalies, or to POC. The interventions were delivered with a mobile phone application and comprised of two sessions per day over 5 consecutive days for a total of 10 sessions. In each session, participants were first exposed to 10 face photographs (anomalous faces in the anomalous faces intervention and POC faces in the POC intervention). Then, participants saw an additional face (anomalous in the anomalous faces intervention and POC in the POC intervention) paired with a story about a morally good deed ostensibly carried about by that person. After reading each moral exemplar story, participants indicated whether they experienced prosocial (e.g., admiration) or antisocial emotions (e.g., anger) and the ease with which they could take the person's perspective. After completing the intervention, participants' implicit biases against anomalous faces and POC were again assessed.

Regardless of the specific intervention variant, congruent implicit biases (i.e., against anomalous faces in the anomalous faces intervention and against POC faces in the POC intervention) were reduced by the NAME intervention. Furthermore, this reduction in congruent implicit biases was evident for both interventions separately, suggesting that the anomalous faces and POC variants of the NAME intervention were independently effective in reducing congruent implicit biases. Importantly, after receiving the anomalous faces intervention, congruent implicit biases against anomalous faces were significantly reduced but incongruent implicit biases against POC were not affected. In contrast, the POC intervention reduced both types of implicit biases and did not have a targeted effect on bias against POC. This result indicates that the POC intervention reduced implicit biases in a nonspecific manner. The anomalous faces intervention, on the other hand, reduced implicit biases against people with anomalous faces in a targeted

fashion that did not affect implicit biases against POC. The most robust reductions in congruent implicit biases occurred in those participants who found it easiest to take the perspectives of the pictured moral exemplars. Taken together, the results of this study suggest that exposure to people with facial anomalies, especially exposure that encourages perspective taking in participants, can reduce negative implicit biases towards people with facial anomalies in a targeted fashion.

The mechanisms underpinning the anomalous-is-bad stereotype are poorly understood. One influential perspective suggests that the anomalous-is-bad stereotype reflects an overgeneralization of an otherwise useful behavioral immune system response that facilitates avoidance of pathogens (Zebrowitz & Montepare, 2008). Since failing to detect pathogen threats can be harmful, this view holds that the behavioral immune system biases behavior towards vigilance against threats to limit the potential for pathogen-induced harm (Haselton et al., 2015; Neuberg et al., 2011; Zebrowitz & Montepare, 2008). The detection of disgust responses to anomalous facial features suggests that anomalous facial features elicit pathogen aversion even when such features do not indicate contagiousness linked to an underlying illness (Ryan et al., 2012; Shanmugarajah et al., 2012; Workman et al., 2021). This view entails the troubling prospect that an aversion to anomalous facial features may reflect a deeply embedded feature of our cognitive architecture, rendering the anomalous-is-bad stereotype difficult to mitigate. Our findings, however, are consistent with recent evidence that the anomalous-is-bad stereotype might not be attributable to pathogen avoidance but is instead learned (Workman et al., 2021, 2022).

Learning that people with visible facial differences were ostensibly engaged in morally praiseworthy behavior was, together with mere exposure to anomalous faces, capable of reducing

negative implicit biases against people with visible facial differences. These findings complement recent evidence that moral vignettes are capable of inspiring prosocial behavior in readers (Han et al., 2022). Behavioral changes were not assessed directly here. The relationship of implicit biases to explicit behavior is not straightforward (e.g., Oswald et al., 2013). Nevertheless, the available evidence finds that, along with harboring negative biases against facial anomalies (Hartung et al., 2019; Jamrozik et al., 2019; Workman et al., 2021), people express their biases in the form of dehumanizing behavior directed against people with anomalous faces (Bull & Stevens, 1981; Houston & Bull, 1994; Rumsey et al., 1982; Workman et al., 2021). The expectation that reduced implicit biases against people with visible facial differences would accompany a reduction in dehumanizing behavior is consistent with these findings, and with a recent report of increased prosocial behavior following exposure to a nearly identical set of stories (Han et al., 2022). Interestingly, a reduction in explicit biases against people with visible differences was not detected after the anomalous faces variant of the NAME intervention. This would suggest that the reductions in negative implicit biases against people with facial anomalies affected by the NAME intervention—and potentially to dehumanizing behavior as well—occur outside of conscious awareness.

A recent study that adapted the same set of moral vignettes found that photographs of typical faces paired with prosocial vignettes were judged more attractive, more confident, and friendlier than typical faces paired with non-moral vignettes (He et al., 2022). Pairing typical faces with antisocial vignettes produced the opposite pattern of effects. Although beyond the scope of the current study, antisocial vignettes may be capable of exacerbating negative implicit biases and dehumanizing behavior directed against people with anomalous faces. This prospect is troubling because visible facial differences are routinely used as visual shorthand in popular

media to indicate underlying moral corruption (Croley et al., 2017; Marion et al., 2018). Perhaps the most flagrant example of this is found in Disney's *The Lion King's*, where the movie's central villain is known as "Scar[2]." This approach to visual storytelling may well have the unintended consequence of aggravating negative attitudes about and behaviors directed against people with anomalous facial features.

Affective and cognitive facets of state empathy were recorded for each participant before the close of each session of the intervention, immediately after participants saw a face photograph paired with a moral vignette ostensibly about an exemplary moral action carried out by that person. Specifically, we recorded frequencies of pro- and antisocial emotions (affective empathy) and ease of perspective taking (cognitive empathy) after participants read each moral vignette and we assessed whether these facets of empathy correlated positively with reductions in negative implicit biases. Almost uniformly, participants reported experiencing prosocial emotions instead of antisocial emotions, with no significant correlations detected between prosocial emotionality and changes to implicit biases. Ease of perspective taking, however, correlated positively with reductions in negative implicit biases. Similar relations were not detected between trait empathy and reductions in bias. Workman and colleagues (2021), on the other hand, presented evidence that the dehumanization of people with visible facial differences is underpinned by trait affective empathy. Although converging on a link between empathy and negative attitudes about people with anomalous faces, these studies differ in which facets of empathy they implicate. This disparity may be attributable to differences in how the studies were designed. Workman and colleagues (2021) used a cross-sectional design to characterize rapidly unfolding brain responses to anomalous faces, which may be particularly sensitive to trait

[2] For more on this: https://www.psychologytoday.com/us/blog/brain-behavior-and-beauty/201908/scarring-your-children-the-lion-king

affective empathy. In contrast, this study used a prospective design that encouraged explicit reflection after exposure to moral vignettes, which may rely more heavily on state cognitive empathy as opposed to trait affective empathy in triggering reductions to negative implicit biases.

We find support for the view that exposure to people with visible facial differences ameliorates negative implicit biases directed against them. The NAME intervention capitalized on two approaches to exposure—mere exposure similar to Zebrowitz and colleagues (2008), and a "humanization" exposure paradigm similar to Majdandžić and colleagues (2012). This study was not designed to disambiguate the differential effects of either kind of exposure, which warrants attention in future research. Regardless of the specific mechanism(s) of exposure underpinning the effect we observed, targeted reductions in negative implicit biases were detected following only the anomalous faces variant of the NAME intervention (mean change in *d*' = -0.29). A smaller reduction in negative implicit biases against POC was detected following the POC intervention (mean change in *d*' = -0.22), which also encompassed a modest reduction in negative implicit biases against anomalous faces that were not targeted by the POC intervention (mean change in *d*' = -0.15). A similar reduction in negative implicit biases against POC was not detected following the anomalous faces intervention (mean change in *d*' = -0.06). These results altogether provide additional evidence that negative implicit biases towards some out-groups, at least—particularly negative implicit biases stemming stem from a lack of exposure to those groups—are malleable.

There is little doubt as to the existence of implicit biases (Jost et al., 2009). There is doubt, however, surrounding some of the methods used to elicit and measure implicit biases. The Implicit Association Test, in particular, has come under scrutiny for failing to predict

discriminatory behavior (Oswald et al., 2013). Nevertheless, the IAT was used in the current study for comparability with previous literature examining negative implicit biases against people with visible facial differences (Changing Faces, 2017; Hartung et al., 2019; Villavisanis et al., 2023; Workman et al., 2021). Furthermore, a rich literature has examined interventions aimed at strengthening and weakening different kinds of implicit biases measured with the IAT (Asgari et al., 2010; Blair, 2002; Blair et al., 2001; Dasgupta & Asgari, 2004; Dasgupta & Greenwald, 2001; Devine et al., 2012; Rudman & Phelan, 2010; Stout et al., 2011; Teachman & Woody, 2003). We are unaware of any such work, however, examining implicit attitudes towards people with visible facial differences. Future work must seek to conceptually replicate our findings using alternative means of eliciting and measuring implicit biases. The Affect Misattribution Procedure (AMP) represents one such alternative—the AMP measures automatic affective responses, in contrast to the categorical associations assessed with the IAT, and has been shown to predict real-world behavior (Lundberg & Payne, 2014; Murphy & Zajonc, 1993; Payne et al., 2005). A dearth of literature, however, has examined the influence of interventions on implicit biases measured with the AMP.

Mobile applications like that used to deliver the NAME intervention are a promising venue for flexibly promoting positive changes to attitudes and behaviors. Many elements of the mobile application could be modified in order to enhance the effects we report, such as the approaches to exposure, session frequencies, and overall intervention duration. An extensive implicit bias training curriculum that incorporates the NAME intervention could bolster the effects described here, although increasing the burden placed on participants could have adverse consequences for session completion rates. Of note, participants in our study completed around 90% of all sessions on average, and the parameters we used were effective in reducing negative

implicit biases against people with facial anomalies in a targeted fashion. A critical objective of future research on the NAME intervention will be to establish, as has been demonstrated for similar interventions (Banaji & Greenwald, 2016), whether the reductions to negative implicit biases against people with visible facial differences that we detected remain stable over time.

Conclusion

This pre-registered study (https://osf.io/b9g6v/) characterized the malleability of negative implicit biases directed against people with visible facial anomalies, like scars and palsies, and identified a candidate mechanism that may facilitate the curbing of such biases. Participants' implicit biases against people with facial differences and people of color (POC) were recorded before and after completing the "Normalizing Anomalies with Mobile Exposure" (NAME) intervention. The intervention itself leveraged exposure to faces, which belonged either to people with visible differences or to POC depending on the type of intervention, in a bid to reduce implicit biases against the people harboring those faces. We predicted that exposure to people with facial anomalies, relative to exposure to POC, would reduce implicit biases against people with such anomalies. The anomalous faces intervention reduced negative implicit biases against anomalous faces but did not change implicit biases against POC. The POC intervention also reduced negative implicit biases, but this effect was observed both for faces that were and were not targeted by the intervention. This finding suggests that, in contrast to the anomalous faces intervention, the effect of the POC intervention was nonspecific and may merely reflect increased familiarity with the implicit bias measure. Interestingly, reductions in implicit biases were positively associated with state cognitive empathy assessed across the intervention. Taken together, the results of this study suggest that exposure to people with facial anomalies,

especially exposure that encourages perspective taking, can reduce implicit biases towards them in a targeted fashion. The NAME intervention is a promising tool for reducing the social burdens shouldered by people who look different, with potential applications generalizing beyond facial differences to include early interventions aimed at reducing implicit biases against other socially penalized out-groups.

Acknowledgements: All authors contributed to the study design. The mobile phone application was developed by NB. Data collection was performed by NB and CIW. Analyses were performed by CIW. The manuscript was drafted by NB and CIW. All authors provided critical revisions and approved the final version of the manuscript. This research was supported by the Penn Center for Human Appearance (awarded to AC) and the National Institute of Dental & Craniofacial Research of the National Institutes of Health (F32DE029407 awarded to CIW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of Interest Statement: The authors report there are no competing interests to declare.

Data Availability Statement: This study was pre-registered (https://osf.io/b9g6v/) and the materials, de-identified data, code, and statistical outputs are publicly available (https://osf.io/vm7d4/).

References

Amodio, D. M., & Cikara, M. (2021). The Social Neuroscience of Prejudice. *Annual Review of Psychology*, *72*(1), 439–469. https://doi.org/10.1146/annurev-psych-010419-050928

Asgari, S., Dasgupta, N., & Gilbert Cote, N. (2010). When Does Contact with Successful Ingroup Members Change Self-Stereotypes? *Social Psychology*, *41*(3), 203–211. https://doi.org/10.1027/1864-9335/a000028

Banaji, M. R., & Greenwald, A. G. (2016). *Blindspot: The Hidden Biases of Good People*. Bantam Books.

Bandura, A. (1969). Social learning of moral judgments. *Journal of Personality and Social Psychology*, *11*(3), 275–279. https://doi.org/10.1037/h0026998

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688. https://doi.org/10.3758/s13428-013-0410-6

Biernoff, S. (2011). The Rhetoric of Disfigurement in First World War Britain. *Social History of Medicine*, *24*(3), 666–685. https://doi.org/10.1093/shm/hkq095

Blair, I. V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828–841. https://doi.org/10.1037/0022-3514.81.5.828

Bull, R., & Stevens, J. (1981). The effects of facial disfigurement on helping behaviour. *Italian Journal of Psychology*, *8*(June), 25–33.

Changing Faces. (2017). *Public Attitudes to Disfigurement in 2017*.

Croley, J. A., Reese, V., & Wagner, R. F. (2017). Dermatologic Features of Classic Movie Villains. *JAMA Dermatology*, *153*(6), 559. https://doi.org/10.1001/jamadermatol.2016.5979

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642–658. https://doi.org/10.1016/j.jesp.2004.02.003

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800–814. https://doi.org/10.1037/0022-3514.81.5.800

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Dey, J. K., Ishii, L. E., Byrne, P. J., Boahene, K. D. O., & Ishii, M. (2015). The Social Penalty of Facial Lesions: New Evidence Supporting High-Quality Reconstruction. *JAMA Facial Plastic Surgery*, *17*(2), 90–96. https://doi.org/10.1001/jamafacial.2014.1131

Doerr, C., Ashby Plant, E., Kunstman, J. W., & Buck, D. (2011). Interactions in Black and White: Racial differences and similarities in response to interracial interactions. *Group Processes & Intergroup Relations*, *14*(1), 31–43. https://doi.org/10.1177/1368430210375250

Erigha, M. (2018). On the margins: black directors and the persistence of racial inequality in twenty-first century Hollywood. *Ethnic and Racial Studies*, *41*(7), 1217–1234.

https://doi.org/10.1080/01419870.2017.1281984

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and Relevant Moral Exemplars Are More Effective than Extraordinary Exemplars in Promoting Voluntary Service Engagement. *Frontiers in Psychology*, *8*(MAR). https://doi.org/10.3389/fpsyg.2017.00283

Han, H., Workman, C. I., May, J., Scholtens, P., Dawson, K. J., Glenn, A. L., & Meindl, P. (2022). Which moral exemplars inspire prosociality? *Philosophical Psychology*, *35*(7), 943–970. https://doi.org/10.1080/09515089.2022.2035343

Hartung, F., Jamrozik, A., Rosen, M. E. E., Aguirre, G., Sarwer, D. B. B., & Chatterjee, A. (2019). Behavioural and Neural Responses to Facial Disfigurement. *Scientific Reports*, *9*(1), 8021. https://doi.org/10.1038/s41598-019-44408-8

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. In *The Handbook of Evolutionary Psychology* (pp. 1–20). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119125563.evpsych241

He, D., Workman, C. I., He, X., & Chatterjee, A. (2022). What is good is beautiful (and what isn't, isn't): How moral character affects perceived facial attractiveness. *Psychology of Aesthetics, Creativity, and the Arts*, 1–38. https://doi.org/10.1037/aca0000454

Hearn, S. L., & Hearn, P. J. (2020). Working With People With Disabilities: An Interactive

Video/Lecture Session for First- and Second-Year Medical Students. *MedEdPORTAL*, *16*(1), 10913. https://doi.org/10.15766/mep_2374-8265.10913

Houston, V., & Bull, R. (1994). Do people avoid sitting next to someone who is facially disfigured? *European Journal of Social Psychology*, *24*(2), 279–284. https://doi.org/10.1002/ejsp.2420240205

Jamrozik, A., Oraa Ali, M., Sarwer, D. B., & Chatterjee, A. (2019). More than skin deep: Judgments of individuals with facial disfigurement. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 117–129. https://doi.org/10.1037/aca0000147

JASP Team. (2019). *JASP (Version 0.11.1)[Computer software]*. https://jasp-stats.org/

Jenkins, A. C., Karashchuk, P., Zhu, L., & Hsu, M. (2018). Predicting human behavior toward members of different social groups. *Proceedings of the National Academy of Sciences*, *115*(39), 9696–9701. https://doi.org/10.1073/pnas.1719452115

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69. https://doi.org/10.1016/j.riob.2009.10.001

Julian, D., & Partridge, J. (2008). *The incidence and prevalence of disfigurement*. https://www.changingfaces.org.uk/wp-content/uploads/2016/03/FE-Campaign-Epidemiology-2-pages.pdf

Knutson, K. M., Krueger, F., Koenigs, M., Hawley, A., Escobedo, J. R., Vasudeva, V., Adolphs, R., & Grafman, J. (2010). Behavioral norms for condensed moral vignettes. *Social Cognitive and Affective Neuroscience*, *5*(4), 378–384. https://doi.org/10.1093/scan/nsq005

Kristjánsson, K. (2006). Emulation and the use of role models in moral education. *Journal of Moral Education*, *35*(1), 37–49. https://doi.org/10.1080/03057240500495278

Kurdi, B., Mann, T. C., Charlesworth, T. E. S., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences*, *116*(13), 5862–5871. https://doi.org/10.1073/pnas.1820240116

Leslie, K. F., Sawning, S., Shaw, M. A., Martin, L. J., Simpson, R. C., Stephens, J. E., & Jones, V. F. (2018). Changes in medical student implicit attitudes following a health equity curricular intervention. *Medical Teacher*, *40*(4), 372–378. https://doi.org/10.1080/0142159X.2017.1403014

Lundberg, K. B., & Payne, B. K. (2014). Decisions among the Undecided: Implicit Attitudes Predict Future Voting Behavior of Undecided Voters. *PLoS ONE*, *9*(1), e85680. https://doi.org/10.1371/journal.pone.0085680

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*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E., & Lamm, C. (2012). The Human Factor: Behavioral and Neural Correlates of Humanized Perception in Moral Decision Making. *PLoS ONE*, *7*(10), e47698. https://doi.org/10.1371/journal.pone.0047698

Marion, T., Reese, V., & Wagner, R. F. (2018). Dermatologic features in good film characters who turn evil: the transformation. *Dermatology Online Journal*, *24*(9), 4. https://doi.org/10.5070/D3249041411

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*,

64(5), 723–739. https://doi.org/10.1037/0022-3514.64.5.723

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience & Biobehavioral Reviews*, *35*(4), 1042–1051. https://doi.org/10.1016/j.neubiorev.2010.08.011

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Rudman, L. A., & Phelan, J. E. (2010). The Effect of Priming Gender Roles on Women's Implicit Gender Beliefs and Career Aspirations. *Social Psychology*, *41*(3), 192–202. https://doi.org/10.1027/1864-9335/a000027

Rumsey, N., Bull, R., & Gahagan, D. (1982). The Effect of Facial Disfigurement on the Proxemic Behavior of the General Public. *Journal of Applied Social Psychology*, *12*(2), 137–150. https://doi.org/10.1111/j.1559-1816.1982.tb00855.x

Ryan, S., Oaten, M., Stevenson, R. J., & Case, T. I. (2012). Facial disfigurement is treated like an infectious disease. *Evolution and Human Behavior*, *33*(6), 639–646. https://doi.org/10.1016/j.evolhumbehav.2012.04.001

Sanderse, W. (2013). The meaning of role modelling in moral and character education. *Journal of Moral Education*, *42*(1), 28–42. https://doi.org/10.1080/03057240.2012.690727

Shanmugarajah, K., Gaind, S., Clarke, A., & Butler, P. E. M. (2012). The role of disgust emotions in the observer response to facial disfigurement. *Body Image*, *9*(4), 455–461.

https://doi.org/10.1016/j.bodyim.2012.05.003

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255–270. https://doi.org/10.1037/a0021385

Sukhera, J., Watling, C. J., & Gonzalez, C. M. (2020). Implicit Bias in Health Professions. *Academic Medicine*, *95*(5), 717–723. https://doi.org/10.1097/ACM.0000000000003173

Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, *112*(1), 100–109. https://doi.org/10.1037/0021-843X.112.1.100

Villavisanis, D. F., Workman, C. I., Zapatero, Z. D., Vu, G. H., Humphries, S., Cho, D. Y., Swanson, J. W., Bartlett, S. P., Chatterjee, A., & Taylor, J. A. (2023). Visual Attention, Bias, and Social Dispositions Toward People with Facial Anomalies: A Prospective Study with Eye-Tracking Technology. *Annals of Plastic Surgery*, 1–23.

White, R. T. (2018). I Dream a World: Black Panther and the Re-Making of Blackness. *New Political Science*, *40*(2), 421–427. https://doi.org/10.1080/07393148.2018.1449286

Workman, C. I., & Chatterjee, A. (2021). The Face Image Meta-Database (fIMDb) & ChatLab Facial Anomaly Database (CFAD): Tools for research on face perception and social stigma. *Methods in Psychology*, *5*, 100063. https://doi.org/10.1016/j.metip.2021.100063

Workman, C. I., Humphries, S., Hartung, F., Aguirre, G. K., Kable, J. W., & Chatterjee, A. (2021). Morality is in the eye of the beholder: the neurocognitive basis of the "anomalous-is-bad" stereotype. *Annals of the New York Academy of Sciences*, *1494*(1), 3–17. https://doi.org/10.1111/nyas.14575

Workman, C. I., Smith, K. M., Apicella, C. L., & Chatterjee, A. (2022). Evidence against the "anomalous-is-bad" stereotype in Hadza hunter gatherers. *Scientific Reports*, *12*(1), 8693. https://doi.org/10.1038/s41598-022-12440-w

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., White, B., & Wieneke, K. (2008). Mere Exposure and Racial Prejudice: Exposure to Other-Race Faces Increases Liking for Strangers of that Race. *Social Cognition*, *26*(3), 259–275. https://doi.org/10.1521/soco.2008.26.3.259

Zeidan, A. J., Khatri, U. G., Aysola, J., Shofer, F. S., Mamtani, M., Scott, K. R., Conlon, L. W., & Lopez, B. L. (2019). Implicit Bias Education and Emergency Medicine Training: Step One? Awareness. *AEM Education and Training*, *3*(1), 81–85. https://doi.org/10.1002/aet2.10124