8-2016

Recognition Training for Faces Across Age Gaps

William Blake Erickson
University of Arkansas, Fayetteville

Recognition Training for Faces Across Age Gaps

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Psychology

by

William Blake Erickson
Henderson State University
Bachelor of Science in Psychology, 2009
University of Arkansas
Master of Arts in Psychology, 2012

August 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

____
Dr. James Michael Lampinen
Dissertation Chair

____
Dr. Denise Beike
Committee Member

____
Dr. Nathan Parks
Committee Member

Abstract

Face recognition is a problem that has theoretical and applied value. However, the fact of facial aging is rarely addressed in research and unmentioned in the major theories of face recognition. Facial aging also has ramifications for missing persons and fugitive cases, confounding attempts by law enforcement to recover these people whose last known images are years or decades out of date. This dissertation reports three studies aimed at measuring baseline age-gap recognition ability and testing various training regimens designed to increase accuracy rates for this unique kind of recognition task.

Acknowledgments

Special thanks are extended to the faculty, staff, and students of the Department of Psychological Science at the University of Arkansas for their support and comradery these six years. Thanks are also extended to Dr. Charlie Frowd at the University of Winchester and Dr. Dawn R. Weatherford at Arkansas State University for their help in carrying out the projects reported here.

Dedication

This manuscript is dedicated to the detectives and investigators who strive to close cold cases involving long-term missing persons and wanted fugitives and also to the relatives and loved ones for who closure may never come.

Table of Contents

I. Introduction

When Boston organized crime boss James "Whitey" Bulger was captured in Santa Monica, California, in 2011, he and his long-term romantic partner had been eluding authorities for almost 20 years. Although they established a residence in California under assumed identities, they traveled widely in their time on the lam (Goodnough, 2011). One can speculate that many law enforcement and transportation officials saw the pair during this time yet failed to recognize them. This is despite the fact that Bulger was a renowned wanted fugitive who did little to change his appearance.

The problem of face recognition across age gaps in instances like the Bulger case has great theoretical and applied value. When people go missing or are on the run as fugitives for many years, facial appearance undergoes predictable changes from natural aging and other changes from lifestyle choices that may affect appearance (Albert, Ricanek, & Peterson, 2007). One way investigators try to overcome the recognition challenge brought by aging is to disseminate forensic age-progressed images designed to approximate an individual's current appearance based on outdated photographs (Taylor, 2001). However, laboratory study of the effectiveness of these images shows they are not judged very similar to their intended targets (Lampinen, Erickson, Frowd, & Mahoney, 2015) and they do not yield greater recognition rates than outdated images alone (Lampinen, Miller, & Dehon, 2012). One reason this may be so is that the visual system already has expertise for faces (e.g., Wallis, 2013), and these images may distract the visual system from making accurate identity judgments. Therefore, an alternate avenue for law enforcement might be to train people to better recognize faces that have aged considerably since their last known appearance. Feedback training has recently shown to be a promising means by which to improve unfamiliar, same-age face matching (e.g., White, Kemp, Jenkins, & Burton, 2014). Such training may also improve face recognition based on outdated study images.

This paper reports a tests of various training regimens designed to improve face recognition across study/test age gaps of 30 or more years. It begins with an overview of the many changes a face undergoes throughout its life. It then reviews the few extant studies specifically investigating this aspect of face recognition and attempts to fit the ability into existing theories. It finally describes new research aimed specifically at improving face recognition accuracy before describing the current studies.

A. Facial Aging

Craniofacial morphogenesis begins four weeks after conception (Gillgrass & Welbury, 2012). At this time, crest cells in the neural fold (i.e., the early analog to the central nervous system) form six arch-shaped structures that eventually become the head and neck. The top two arches transform into the muscular, arterial, and skeletal superstructure of a recognizably human face by week 10. Cartilage and soft membranes ossify into facial bones so that by birth the brain is protected when it exits the birth canal.

Development continues after birth in a series of anabolic growth processes (Ramsey, Marcheva, Kohsaka, & Bass, 2007). Importantly, the brain continues to grow such that it reaches 90% of its adult size by five years (Gilgrass & Welbury, 2012). The skull grows during this time to compensate, forcing facial bones to grow or recede along their edges in processes called deposition and resorption, respectively. The bones eventually fuse together during adolescence. As the rest of the body gradually catches up to the brain's relatively larger size, the skull elongates vertically and the jaw moves forward, taking a cardioid shape by 20 years (Enlow & Hans, 1996; Pittenger & Shaw, 1975).

Further development from 20 years onward is the primary concern of the current study. Whereas the growth up to this time incurs extreme changes in craniofacial shape and size, the final adult proportions remain relatively constant for many decades. Subtle textural changes do develop at the corners of the mouth and eyes and on ridges across the forehead due to frequent hyper-dynamic facial expressions (Albert et al., 2007). Complexion becomes less even-toned as effects of ultraviolet light exposure accumulate over time. After 50 years, skin loses elasticity, cartilage in the nose and ears continues to grow, and the jaw shortens due to gum deterioration and tooth loss. The natural progression of these events varies somewhat among individuals but the sequence is the same for everyone. Additionally, lifestyle factors such as drug use, sleeplessness, stress, weight, and extended time in direct sunlight speed the progression of these changes, which is most apparent in studies of identical twins that live apart (Guyuron et al., 2009).

Changes related to aging like those described above constitute holistic changes in facial appearance. In other words, they affect a face's shape and texture while preserving the configuration and location of individual features. Judgment of facial age itself is holistic, biased toward the age of a face's bottom half because this area contains the greatest age-related variability (Hole & George, 2011). As described earlier, face perception is a holistic process susceptible to minute disruptions of featural configuration. That being said, some holistic changes are actually beneficial. Caricatures are recognized more quickly and more accurately than veridical representations of faces in part because they exaggerate the holistic aspects of faces that are already unique (Rhodes & McLean, 1992). An older face takes on a caricatured appearance of its younger counterpart as cartilage continues to accumulate in the ears and nose and lines become exaggerated (e.g., Hancock & Little, 2011). Therefore, a reasonable prediction might have older adult test faces producing higher recognition rates than younger test faces when young faces are studied. Although extant research has shown that this is not the case, recognition across age gaps remains more accurate than chance.

B. Recognizing Faces across Age Gaps

Cases such as that of Whitey Bulger by their very nature confound investigators. Age-related appearance changes complicate recovery efforts whether they occur within adulthood or from childhood into adulthood. Although forensic artists may attempt to incorporate general knowledge about craniofacial morphology and familial aging patters into age-progressed images, these images remain educated guesses that may vary widely from artist to artist (Lampinen, et al., 2015). Therefore, they may lead investigators and the public to look for the wrong individuals. Moreover, the human face recognition system adeptly compensates for age-related changes already.

Face Perception and Recognition. Little research has examined face recognition as targets age, which is unusual given that aging is a natural process. The dearth of examinations may be due to the fact that the human visual system can compensate for age-related facial appearance changes rather well and is therefore of little interest to investigators. Seamon (1982) termed this ability "bidirectional dynamic facial recognition" (p. 370). Using several recognition and matching paradigms, he showed that people are able to match even unfamiliar faces across age gaps well above chance after short periods of aging. In his first experiment, undergraduates studied photographs of faculty members from their university's 1974 yearbook for ten seconds each. Recognition memory was tested using a set of photographs that contained either the same 1974 photographs mixed with images of foil individuals or photographs of the same faculty members from the school's 1966 yearbook mixed with foil individuals. Recognition rates were higher when test age matched study age, but the average *d'* score of the age gap group averaged 1.10, which is still above chance. A second experiment manipulated familiarity by giving participants 14 photographs to study for five minutes, after which 28 images (half from the study phase) were studied for ten seconds each as in the first experiment. Same-age photographs and more familiar photographs were more accurately recognized than younger and unfamiliar photographs. A third experiment was carried out where the 1966 photographs were used for study and the 1974 photographs were used in testing, and similarly to the first two experiments, repeated images increased recognition over different-age images and familiar faces were better recognized than unfamiliar ones. The fourth experiment tested whether facial recognition of different-aged faces can occur incidentally rather than intentionally. To do this, the author had participants perform a sorting task with the 1966 photographs where decided which instructors would be "hard graders" or "better teachers" based on facial appearance, followed by a surprise recognition test including either 1966 or 1974 tests. Results were similar to previous studies, extending the effect to incidental learning. The fifth study did not measure recognition, but rather the ability to match face images of the same individuals at different ages. Additionally, its stimulus images were children rather than adults. Pictures of individuals at their teenage, prepubescent, and infant photographs were given to participants who were tasked with matching these images to adult (~ 20 years) images of the same individuals. Results revealed a clear, gradual decrease where 95% of adult/teenage pairings were correct, followed by 84.74% adult/prepubescent pairings, and ending with 55.60% young adult/infant pairings (see Figure 1). The remarkable aspect of this result is that chance level for pairing correctly was 15.48%.

Bruck, Cavanagh, and Ceci (1991) conducted a field study of face recognition across age gaps utilizing images taken from yearbooks and measuring name-matching ability of classmates from a 25th year high school reunion. These and current photographs of the individuals were mixed with foils and included in test booklets mailed to participants, who were told to name five outdated high school photographs and then match them to one of ten photographs of people in their mid-40s, of which five were the same individuals and five were visually similar individuals. Participants also provided information about their familiarity with the high school individuals as well as how recently in years they had seen them. A control group comprised of individuals from a different country who could not have known any people in the test or study images was also tested. Classmates who had seen the individual within the past 17 years were excluded from analyses. The chance level of matching images was 10%, and 49% of classmates correctly matched high school images to the older images, and 21% provided names, of which 71% were correct. More impressively, 33% of individuals in the unfamiliar control group correctly matched the younger and older images. This indicates that base perceptual information was sufficient to guide recognition, and not just familiarity with the targets. One limitation of the study, which Bruck et al. concede, is that the return rate for the classmate group was 48%, and those most familiar with the images were more likely to send the test booklets back. Therefore, the matching and naming data might be overestimated in the classmate group, and true rates could be much closer to the unfamiliar group. Another limitation of is that the degree of physical change apparent in the facial images from one age to another varied greatly, particularly in facial hair, hairstyle, and weight. The authors did not attempt to quantify these changes in any way, and therefore could not correlate image differences with accuracy rates. However, the study demonstrates the longevity of facial memory and also provides a more ecologically valid estimate of cross-age gap matching accuracy for familiar and unfamiliar faces.

Missing Persons Research. Recent investigations of the efficacy of age-progressed images have incidentally revealed that face recognition is robust to age-related appearance changes. Law enforcement renders these images ostensibly to increase recognition rates of targets in long-term missing persons and fugitive cases. In the earliest study, Lampinen, Arnal, Adams, Courtney, & Hicks (2012) commissioned age-progressed images of volunteers' childhood images. Participants either studied age progressions, outdated photographs, or current photographs before being given a recognition test using current photographs of targets and foils. Current study photos yielded highest recognition rates, but age progressions and outdated images did not differ. Lampinen, Miller, & Dehon (2012) examined the efficacy of these images in prospective and retrospective person memory paradigms. Such paradigms attempt to simulate the search for missing persons in an ecologically valid fashion based on event-based prospective memory. In other words, participants are given a prospective task (e.g., "keep an eye out" for this individual") and an ongoing task (e.g., sort these groups of individuals into two teams of equal males and females) during which the prospective cue is presented. In the prospective person memory task, recognition rates based on outdated study images were marginally more accurate than recognition based on age-progressed images. In the more traditional recognition memory task, outdated and age-progressed recognition rates did not differ. Furthermore, unpublished data by Erickson, Lampinen, Frowd, & Mahoney (2013) replicated the outdated photo advantage in a large design containing many different age-progressed images from multiple professional forensic artists (see Figure 2). Although outdated study images did not yield reliably higher recognition rates than age progressed images, a difference in discriminability of nearly 10% equates to hundreds of real cases (NCMEC, 2016).

Taken together, these studies demonstrate three major points concerning the human facial recognition system's ability to adapt to the changes that age brings to human faces. First, memory for faces lasts a very long time, and this memory is strongly mediated by familiarity. Second, people are able to make accurate identity judgments based on basic perceptual information from an unfamiliar aged face, such as its pose. Third, although few studies have systematically examined this specific ability, they show it has remained stable across three different decades. In spite of these points, face recognition theories have failed to address how the visual system compensates for age-related appearance changes.

C. Recognizing Faces

The most-cited theoretical account of recognizing faces is Bruce and Young's (1986) model for recognizing familiar faces. Like most cognitive models, it begins with an input signal – in this case, a face – that is then decomposed into more primitive "structural codes" robust to viewpoint and lighting changes. These in turn activate "facial recognition units" that represent stored familiar faces. The final steps, encompassing the actual recognition, are activations of identity nodes holding semantic information about the person and the eventual generation of a name[1]. Although such a framework neatly outlines the recognition scenario people experience on a daily basis, it leaves out the effect of a long-term gap in seeing familiar individuals. Granted, a person seen often usually retains his general facial appearance between viewings. However, it is also common to experience a moment of confusion when a person we know changes appearance in a simple way, such as a subtle alteration to hairstyle. This is in spite of the internal features' remaining unchanged. Therefore, the myriad changes brought on by facial

[1] The model also incorporates stages for expression recognition, visual cues for speech perception, and gaze direction, but these are not of theoretical pertinence to the current study.

aging over the span of years and decades should be especially taxing on the recognition system. However, as discussed above, humans have little trouble with this task.

People primarily recognize faces in a holistic fashion (Farah, Wilson, Drain, & Tanaka, 1998; Richler & Gauthier, 2014). That is to say, qualities of a face that contribute to its surface texture and overall configuration are used as recognition criteria rather than separate individual features. In this way, a face is a gestalt greater than any of its parts. However, the degree to which holistic processing occurs can vary depending on who is being perceived. For example, unfamiliar faces (Megreya & Burton, 2006), other-race faces (Michel et al., 2006), and faces belonging to other social categories (Hehman, Mania, & Gaertner, 2009) are all recognized less accurately than familiar faces within one's own race or social category. Furthermore, evidence supports the possibility that they are processed at a level where individual internal features receive more attention than the whole faces (Tanaka & Farah, 1993).

Holistic processing develops as a byproduct of facial expertise (Gauthier & Tarr, 1997; c.f. Farah et al., 1998). As far as the visual system is concerned, a face is just another object. However, since humans spend so much time looking at faces, holistic processing arises to aid the visual system in distinguishing among the hundreds of faces encountered over a lifetime, all of which possess the same configuration. This is in contrast to how other natural objects such as rocks or trees can vary widely in shape. Therefore, it is quite reasonable that people have greater expertise for familiar, own-group, and own-race faces as described above, which is why they process these faces at a more holistically than other faces.

Another popular theoretical account of facial processing is Valentine's (1991) "face space". Face space refers to a multidimensional framework in which neural representations of faces are stored. Each dimension represents a biometric measurement of the face, such as the length of the nose or distance between the eyes. Faces that appear similar to one another will be closer together in this space than faces that are dissimilar to one another. Moreover, every visual system's face space has a different architecture built through experience the myriad of faces observed over a lifetime. Consequentially, face types for which the viewer has most experience (e.g., from one's own race) are perceived as more distinct and far apart in the space, whereas other face types cluster together more densely and are perceived as more similar. The viewer must learn the unique type of variance that defines a novel face type. It should be pointed out here that distinctiveness as it relates to face space only reflects the viewer's perception and not necessarily actual biometric properties of the faces. The face space framework accounts for a number of holistic face recognition findings, and may be valuable in determining how faces can be recognized across age gaps.

Valentine's face space has been supported by recent advances in computer vision applications of automatic face recognition. The most common analog for human recognition ability is the principle-components analysis (PCA) system called "eigenfaces" (Turk & Pentland, 1991). PCA is normally used as a statistical technique to determine what discreet clusters of variability contribute to a dataset's total variance (Tabachnick & Fidell, 2001). Since a computer processes faces as numeric pixel data, it examines sets of face images in the same way it would analyze any other data. To perform a PCA on a set of 2-dimensional face images, computer algorithms first break down input images into one-dimensional vectors of grayscale brightness values. The PCA then uses these input vectors as individual "subjects" to calculate the components that contribute to the entire dataset's overall variance. Since facial images are analyzed in this case, the variability in each component directly reflects discrete variations in facial appearance throughout the library of input images. Each component includes an eigenvector, which can be reconstructed into a 2D image called an “eigenface”, which is a ghostly representation of the facial primitive that highlights areas where the most variance is being captured in the corresponding component. A face from the original corpus can then be “recognized” later by comparing a new input face to faces reconstructed from the learned eigenfaces. Typically, the most variance is due to overall face shape as visible in early eigenfaces, and the least variance is due to idiosyncratic differences in internal facial features and face texture. Eigenfaces are analogous to the multiple dimensions found in Valentine’s (1991) face space, as they vary holistic features. Additionally, a large image set featuring exemplar faces from both sexes and many different races would also output early shape-based eigenfaces that reflect variance based on race and sex (Abdi, Valentin, & Edelman, 1998).

Given the general expertise the visual system has for faces, its ability to compensate for age-related appearance changes might be expected. To date, no theoretical account, cognitive or computational, has addressed facial aging. Forensic art experts refer to the “life-long look”, or invariant features across time, but stop short of identifying them or speculating their impact on face perception (Sadler, 1986). Seamon (1980) speculates that internal features, such as the mouth and eyes, naturally attract the most attention of viewers and remain fairly consistent in appearance and configuration regardless of age. This possibility was especially tested in his fifth experiment, where key features must have attracted enough attention for the high rates of correct pairings to have been observed. It stands to reason that a PCA-based face space model would capture fine-grained age-related changes such as wrinkles or blotchy complexion in later eigenfaces, but shape changes incurred by anabolic and late-in-life catabolic growth would be captured by earlier eigenfaces. Although these are holistic features, internal features (found in the middle region of eigenfaces) are much less likely to change position on a human face. This leaves open the possibility that, in the wake of large holistic changes, internal features are used as diagnostic cues to a person's identity after a long period of aging. Whatever the case, systematic training and experience are the only way to increase facial aging expertise and recognition rates.

D. Improving Face Recognition

Interest in attempts at improving face recognition has increased in the past decade, motivated by the same security and law enforcement concerns that inspired the current study. These investigations coincided with the development of automated software-based face recognition systems. O'Toole et al. (2007) illuminated the shortcomings of human face recognition compared to computerized image processing algorithms, comparing performance of human participants to six different algorithms. The authors created "easy" face pairs defined as far from one another in a PCA face space. Specifically, these faces contained different feature-level features but minimal holistic differences. "Hard" pairs were relatively close to one another. Algorithms make decisions almost instantly, but humans were given from less than a second to unlimited time to make their identity matching decisions from trial to trial. Humans and algorithms performed equivalently when face pairs were easy or when humans were given more than two seconds to decide, but algorithms outperformed humans on difficult pairs and when humans were given less than two seconds. With these results, O'Toole et al. call attention to a major applied problem of face recognition: Namely, machines are more accurate than human observers particularly when task difficulty increases. Software engineers examining the problem of facial aging, which they term "probe-gallery currency", have nonetheless found that machines' error rates increase as study and test ages grow more distal (Ricanek & Tesafaye, 2006). Another limitation is that machines cannot be everywhere that human authorities are.

Therefore, researchers have examined ways in which face recognition may be improved via training.

All faces share a first-order structure, meaning that they have a predictable configuration such that eyes are positioned above the nose, which is positioned above the mouth. For this reason, face perception and recognition are finely tuned for upright faces, and these processes are greatly disrupted when faces are inverted. The visual system can be trained to compensate for novel presentations such as inversion, improving recognition for inverted, upright, and scrambled faces when individuals are trained specifically for each (Hussain, Sekuler, & Bennett, 2009). Importantly, this training is transferable to new, unfamiliar faces.

Further training studies have examined more natural changes to faces. Such changes constitute within-person variability, a relatively understudied area in face perception. Faces undergo many changes in expression, luminance, pose, and of interest to the current study, age. Such changes have little impact on familiar faces, but unfamiliar faces are particularly vulnerable to these variations. Jenkins, White, Montfort, & Burton (2011) had British and Dutch participants sort 40 photographs of two Dutch celebrities by identity without telling the participants how many identities were in the set. British participants sorted the photographs into an average of 7.5 identities, whereas Dutch participants performed almost perfectly. This illustrates the difficulty of integrating dissimilar appearances of unfamiliar people.

To combat recognition errors caused by within-person variations, White, Kemp, Jenkins, & Burton (2014) designed a training regimen based on the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010). The GFMT consists of matching and mismatching face pairs with matching pairs taken on the same day but with different cameras. Participants took an abbreviated form of the GFMT containing 40 trials (half matching, half mismatching). After each trial, they either received feedback about their accuracy or did not. Feedback training increased accuracy from 82% to 92% as trials progressed, whereas proceeding through the task with no feedback showed no improvement. Moreover, training benefitted those participants with the lowest face matching aptitudes (i.e., one standard deviation below the average matching rate) most. Dowsett and Burton (2014) also examined face matching accuracy when judges made decisions in pairs, mirroring common real-world face identity verification scenarios. Across three experiments, pairs consistently outperformed individual judges (around 80% vs. around 72% respectively), and training in pairs improved matching accuracy of the less apt member when later tested alone.

To date, no training studies have included age gap between study and test faces as a factor. However, there are good reasons to expect training to improve face recognition of this kind. As described earlier, although errors increase as study and test face ages increase, matching and recognition nonetheless exceed chance. This is likely because faces of the same individual at different ages are nearer each other in face space than faces of different individuals. Dakin and Origie (2009) conceptualize this as an "identity trajectory" along which faces travel in face space compared to an average face. Aging, then, might merely nudge faces along their identity trajectories with time. Faces do not age randomly, so every face will travel along similar age-related feature vectors like those described above. Training (with feedback or not) could produce perceptual adaptation to these changes. Exposure to older adult faces may make the age transformations easier for trainees to perceive and compensate for. Feedback training specifically might aid trainees in adopting perceptual strategies when making identity judgments across age gaps. These are possibilities explored in the current project.

E. The Current Study

The problem of recognizing faces that have aged since their last viewing poses a challenge to theory as well as relevant forensic scenarios. The current study examines training regimens designed to improve unfamiliar face recognition across age gaps. Importantly, it seeks to answer the question of whether training people to recognize faces across age gaps is necessary in itself or whether training with same age faces suffices to improve this ability. To answer this, some participants will undergo a training regimen with same age faces at study and test while other participants undergo training with different age faces at study and test. For the sake of completeness, some participants will study young faces and test on older faces while some participants will study older faces and test younger faces. Given that faces naturally age younger to older, recognition in this condition will likely be more accurate than the opposite direction. However, in a forensic scenario, officers may switch between studying outdated photographs to recognize aged suspects and comparing older faces to outdated photographs. Regardless of direction, aging nonetheless represents a holistic change to facial appearance. In keeping with previous studies, training either includes feedback such that participants are informed trial-by-trial whether their judgments are correct or does not. Based on the research reviewed above, the following hypotheses were formulated:

1. Recognition across age gaps will be less accurate than recognition within the same age.
2. Recognition from the younger to older direction will be more accurate than recognition from the older to younger direction.
3. Training will increase accuracy of posttest recognition judgments compared to pretest.
4. Training with feedback will increase accuracy of recognition judgments greater than training without feedback.

5. Training across age gaps will increase recognition judgment accuracy greater than same-age training.
6. Training trial direction (i.e., progression and regression) will preferentially improve accuracy of respective posttest trial directions compared to pretest.

II. Experiment 1

A. Method

Participants

Two hundred thirty-seven college undergraduates across two universities participated in this study in exchange for credit toward a research participation requirement in their introductory psychology classes. Participants were recruited from the University of Arkansas and Arkansas State University. Detailed information about the participants can be found in Table 1. Because stimuli consisted of images of famous actors and musicians, participants were selected from a pool who responded to prescreening questions indicating lack of familiarity with these individuals on Likert-type scales (see Appendix A for these questions).

Materials

Images of 180 actors and musicians (half male, all Caucasian) were found and saved from Google Images searches. To reduce likelihood that participants would be familiar with the identity of these faces, all faces belonged to actors and musicians most popular at or before the 1990s (see Appendix B for a list of face identities used). Identities were selected from lists of winners of various entertainment awards. Face images were found for each celebrity by entering key words including name and the desired age (e.g., "Clark Gable age 20") or, when specific ages could not be found or verified, general age range (e.g., "Joan Jett young"). Final images were selected only if faces were mostly front-facing and contained no occlusions. Two images each at around age 20 (young adulthood) and age 50 or above (older adulthood) were retrieved for each identity, equaling four images total per individual. Identities were paired for test and training trials based on their perceptual similarity to one another.

Regardless of images' native resolution and color profile, all were cropped to prominently feature the face, converted to greyscale JPEGs, normalized for contrast, and resized to 250x350 pixels using Adobe Photoshop to ensure standardized presentation. Stimuli were presented on Dell Optiplex desktop computers with 60 Hz monitors displaying 1366x768 and 1440x900 resolutions. Faces occupied .72 to 6.54 degrees of vertical visual angle (M = 3.01, SD = .91) with participants seated approximately 24 inches from monitors. Experimental session files were created using E-Prime stimulus presentation software, which also recorded data.

Design and Procedure

The study employed a 2 (Test Trial Type: Progression, Regression) x 2 (Feedback: Given vs Not) x 4 (Training Type: Progression, Regression, Same Young, Same Old) mixed design, where test trial was a within-subjects variable and feedback and training type were between-subjects variables. Participants underwent sessions alone or in groups of up to five. Participants viewed series of trials where they were instructed to study faces for 2s before automatically being shown a Gaussian mask for .5 seconds. Next they were instructed to choose the previously studied identity from a self-paced 2-item forced choice test. Upon selecting a face, a new study trial automatically began. If a trial included feedback, feedback would appear after the test in the center of the screen with a cumulative percentage of successful trials below it. Positive feedback displayed green text and negative feedback displayed red text. Study faces and mask displayed at the center of the screen, and test faces appeared side-by-side while vertically centered. Participants chose a face by pressing the "Q" key for the left face and the "P" key for the right face, and left/right positioning was randomized. This general process repeated throughout the duration of the experiment session (see Figure 3 for an example of the basic sequence).

Sessions contained three phases. In Phase 1, participants took a 20-trial pretest. Half of these trials showed age progression (i.e., study younger adult image, test two older adult images) and half showed age regression (i.e., study older image, test two younger). Half the faces were male and half were female. This pretest established participants' baseline face recognition ability. Phase 2 employed a similarly designed 50-trial training regimen with different faces. Participants were randomly assigned to one of two training regimens: one where they either received feedback after each test trial or one where they did not (making the no-feedback condition functionally identical to the pretest). Training was also further divided into a specific age categories that lasted the duration of training: progression, regression, same-age younger (i.e., study young age, test two young faces), or same-age older trials. Phase 3 was a 20-trial posttest using different identities than pretest or training. Figure 4 displays a schematic diagram illustrating the design of study sessions. Pretest and posttest were counterbalanced in two ways, making four discrete counterbalancing conditions. First, half of participants experienced one set of identities as the pretest and the other set of identities as the posttest. Second, half of the identity pairs were used as regression trials and the rest were progression trials in one version of the test, and the other version switched the aging direction. All test trials were presented at random within their phase.

B. Results

The main dependent variables of interest are proportions of accurate recognitions during training and change in accuracy calculated as the difference in hits between posttest and pretest. The former serves to replicate previous research by comparing face recognition accuracy across age gaps to recognition within an age range. The latter tests the efficacy of the various training regimens employed in the current manuscript, comparing stimulus age and feedback presence.

The first analyses test potential effects on overall accuracy unrelated to the main hypotheses. To examine this, overall average accuracy scores were computed across training and test trials. No effects of counterbalancing group were found, with accuracy ranging from 68% to 70%. Collection site also had no effect, averaging 68 and 71%. Male and female participants' accuracies were equivalent, and participant age was uncorrelated with accuracy.

One limitation that arose during stimulus image collection was that the facial ages could not be verified. Therefore, an experiment was constructed where each of the 560 facial images were presented sequentially to 30 participants in a random order. Participants were tasked with estimating the ages of each facial identity to the nearest year. Older images (M = 54.52, SD = 4.31) were estimated to be older than younger images (M = 29.49, SD = 4.96), $t(16) = 11.46$, $p < .001$, *Cohen's d* = 5.39. This verifies that older adult images were in fact perceived to be older than younger adult images of the same identities.

Another consideration worth examining is item difficulty. Appendix A contains not only pairing identities but accuracy rates within each age range and direction. Sixteen of 380 unique pairings yielded recognition rates less than 50% (i.e., less than chance), and 19 pairings yielded accuracy above 90%. Table 2 shows detailed information about each age group and direction collapsed across training and test trials. Of note are skewness and kurtosis values, which fall within moderate normality thresholds (Curran, West, & Finch, 1996). Having demonstrated stable, normal accuracy scores among age ranges and directions, the primary dependent measures will be analyzed below.

Primary Analyses. The first analysis examined accuracy within training trials only. To determine possible fatigue effects, trials were organized into five sequential bins of 10 trials each, creating a five level within-subjects factor to add to the aging direction and feedback between-subjects factors. A Huynh-Feldt corrected within-subjects test found no effect of bins on accuracy, and thus no fatigue effects over time. Likewise, feedback had no effect on accuracy. A main effect of aging direction did manifest, $F(3, 229) = 115.45$, $p < .001$, $n^2_p = .602$ (see Figure 5). Tukey's HSD tests revealed the effect was driven by same age older faces again providing greater recognition than remaining ages, p's < .001. Same age younger faces yielded greater recognition than progression and regression directions, p's < .001, and the latter two were not different from one another.

The main analysis examining change from pretest to posttest also used the same design as before. Progression test trials tended to improve more after training than regression test trials, $F(1, 156) = 22.14$, $p < .001$, $n^2_p = .088$ (see Figure 6). Training age direction yielded no effect, but experiential training led to more accurate recognition than feedback training, $F(1, 229) = 3.93$, $p = .049$, $n^2_p = .017$. No interactions were observed.

Cross-Race Concerns. The own-race bias refers to visual systems' optimal processing of own-race or own-ethnicity faces (Meissner & Brigham, 2001). All stimulus images showed Caucasian celebrities. Thirty-one percent of the sample consisted of non-Caucasians, which is too few to make inferential comparisons of participant race yet enough to influence the overall data patterns. The following analyses are identical to the previous set except they included only Caucasian participants.

The first analysis examined accuracy within the training trials, revealing a significant effect of aging direction, $F(3, 156) = 83.41$, $p < .001$, $\eta^2_p = .616$ (see Figure 7). Tukey's HSD tests revealed the effect was driven by same age older faces again providing greater recognition than remaining ages, p's < .001. Same age younger faces yielded greater recognition than progression and regression conditions, p's < .001, and the latter two were again not significantly different from one another. Feedback and training bins produced no reliable differences in accuracy, and no interactions were observed.

The main analysis examining change from pretest to posttest also used the same design as before. Progression test trials tended to improve more than regression test trials, $F(1, 156) = 13.07$, $p < .001$, $n^2_p = .077$ (see Figure 8). Training age direction this time yielded a main effect, $F(3, 156) = 3.37$, $p = .02$, $n^2_p = .061$. Tukey's HSD tests revealed that the regression and same age older training regimens improved posttest scores more than progression training. All remaining comparisons and interactions were nonsignificant.

Exploratory Analyses of Gender. After initial analyses were conducted, a further test was conducted to determine if target sex influenced recognition accuracy. A Huynh-Feldt corrected repeated measures ANOVA analyzing training trials showed that participants were more accurate recognizing female faces (M = 73.72%, $SD = .13$) than male faces (M = 70.43%, $SD = .13$), $F(1, 229) = 16.05$, $p < .001$, $\eta^2_p = .065$. Target sex interacted with direction, $F(3, 229) = 3.61$, $p = .01$, $\eta^2_p = .045$, driven by male faces being less accurately recognized in the same age younger ($p < .001$) and progression ($p = .01$) trials. An examination of training found no main effect of target sex, but target sex interacted with trial type, $F(1, 229) = 19.00$, $p < .001$, $\eta^2_p = .08$. Simple effects tests revealed that male target accuracy was greater in progression trials ($p = .002$) but female target accuracy was greater in the regression trials ($p = .009$), a result that will be explored in the discussion.

C. Discussion

Experiment 1 tested several training regimens designed to improve recognition of faces when study images portray a different age than test images. In addition to examining whether recognition training across age gaps improves age gap face recognition more than same-age face training, it tested whether trial-by-trial feedback affects post-test recognition differently than experiential training. Taken together, the current results yielded some important findings.

First, same age recognition was found to be more accurate than age gap recognition. Collapsed across feedback and aging group, same age recognition averaged about 80% accuracy, whereas age gap recognition averaged about 64% accuracy. This demonstrates the proof of concept that face recognition within an age range is more accurate than recognition across age gaps, and also supports Hypothesis 1. Feedback, however, did not affect accuracy nor interact with trial bin to show a cumulative improvement over time, which has been observed elsewhere in the literature.

Concerning posttests, progression trials averaged a 5% increase in accuracy regardless of training, whereas regression trials averaged a 5% decrease in accuracy. This was only observed at the within subjects level of test trial type, and not the between subjects level of training trial type. Therefore, it cannot be taken as support for Hypothesis 2. Counter to hypotheses, feedback training yielded a net 2% decrease in posttest accuracy compared to pretest, whereas experiential training yielded a 2% increase. Before any major assumptions can be made, it is worth pointing out one particular reason why effects of training may be difficult to interpret.

One limitation to this experiment is that face stimuli consisted of celebrity identities. These images were chosen partially out of convenience but primarily because their “in the wild” (e.g., Huang, Ramesh, Berg, & Learned-Miller, 2007) image-level variability more closely matches what investigators would encounter in field settings. Although participants were prescreened for lack of familiarity with celebrity faces, they were nonetheless given the opportunity to disclose any familiarity they may have had with stimulus faces and film, television, and music performers more generally. Table 3 shows the number of faces participants indicated they found familiar. Alarmingly, only 30% of participants indicated no familiarity with any stimulus faces despite our prescreening procedures. When given the opportunity to name faces they found familiar, 21 participants indicated they could not name the individuals but could otherwise provide descriptive or identifying information (e.g., "Rizzo from Grease", etc.), and 120 could name at least one face. Table 4 shows responses for general familiarity for classic film, television, and music performers. Again, less than half of participants indicated they had no familiarity with these classes of individuals.

Relative familiarity with various stimulus faces may explain some of the current experiment's results. In particular, our younger adult sample would likely be more familiar with older appearances of many of the older celebrities (e.g., Ozzy Osbourne, Betty White, and Clark Gable) rather than their younger appearances. This might explain the advantage in recognizing Same Older trials over Same Younger trials during training and the progression test trial advantage. Likewise, most of the celebrities participants could freely name were female, which would explain the 3% boost in recognizing female faces over male faces during training. For these reasons, Experiment 2 aimed to replicate Experiment 1 while eliminating the possibility that participants would be familiar with the facial stimuli.

III. Experiment 2

Results in Experiment 1 revealed general promise for face training regimens that include age gaps. However, due to widespread familiarity with facial stimuli and the capricious image-level variability of photographs, the actual benefits of training are difficult to interpret. So, for the current experiment, artificial faces were constructed using EvoFIT, a forensic composite construction program that breeds novel faces using genetic algorithms (Frowd, Hancock, & Carson, 2004). Importantly, these faces are generated from photo databases of real human faces input into a principal-components analysis (PCA) of pixel brightness values. Novel faces can be generated by randomly weighting and combining eigenfaces derived from a specific database of individuals who share general demographic similarities (e.g., Males 17-23 years). After construction, novel faces can be transformed by weighting them toward eigenvectors that correspond to subsets of the library (see Frowd et al., 2006, for a complete description of this process). Conceptually, this is equivalent to making a holistic change in appearance captured by the given subset. In the current case, the aging subscale was used to distort the novel images to add several decades of facial appearance (see Figure 9).

Experiment 2 attempted to replicate Experiment 1 but with greater precision in the form of using novel faces generated and aged with EvoFIT. These faces cannot be recognizable because they are artificial. Therefore, they may provide a clearer assessment of the general hypotheses. This study also differed from the Experiment 1 in that it did not use a "same age" condition, because an image of a novel identity cannot be generated in EvoFIT more than once.

A. Method

Participants. One hundred forty-eight college undergraduates across two universities participated in this study in exchange for credit toward a research participation requirement in their introductory psychology classes. Detailed information about the participants can be found in Table 1. Because novel faces were generated and aged for this experiment, no familiarity prescreening was necessary.

Materials. One hundred eighty facial images were generated using EvoFIT using a random combination of shape and texture vectors. Half of these originated from the Caucasian male age 17-20 years database and half from the related female database. After faces were generated and saved, they were imported into EvoFIT's holistic aging tool. The tool's user interface takes the form of a sliding bar that the operator can drag to the right (i.e., increasing apparent age) or left (i.e., decreasing age). The extremes correspond to one standard deviation away from the initial image along the holistic scale. In the case of age this corresponds to roughly 15 years of age-related appearance changes. The slider can also be reset after holistic changes have been saved so that the operator can apply further changes in the same direction along the scale. So, each face was generated from the younger adult male or female database, loaded into the aging holistic tool, and transformed with two extreme applications of forward aging. Resultant older adult images appear decades older than their younger counterparts. Additionally, faces were not provided with hair to prevent any capricious effects of hairstyle change.

Resultant images were greyscale JPEGs sized 180x240 pixels. Stimuli were presented on Dell Optiplex desktop computers with 60 Hz monitors displaying 1366x768 and 1440x900 resolutions. Facial height is constant for EvoFIT faces, with female faces occupying 1.79 degrees of vertical visual angle and male faces occupying 2.39 degrees of vertical visual angle. Participants sat approximately 24 inches from monitors. Experimental session files were created using E-Prime stimulus presentation software, which also recorded data.

Design and Procedure. The study employed a 2 (Test Trial Type: Progression, Regression) x 2 (Feedback: Given vs Not) x 2 (Training Type: Progression or Regression) mixed design, where test trial was a within-subjects variable and feedback and training type were between-subjects variables. Otherwise, procedures were identical to Experiment 1.

B. Results and Discussion

The main dependent variables of interest are the same as those from Experiment 1. Again, no effects of counterbalancing group on overall accuracy were found, with accuracy ranging from 65% to 70%. Collection site also had no effect, averaging 67 and 69%. Age and gender were unrelated to overall accuracy.

Given that facial stimuli were artificial, an experiment was constructed to validate appearance based on age. Each of the 260 facial images were presented sequentially to 25 participants in a random order. Faces aged with holistic tools (i.e., the older adult faces; M = 42.11, SD = 5.69) were judged significantly older than younger images (M = 24.95, SD = 5.64), $t(23) = 7.52$, $p < .001$. This verifies that holistic tools produced faces that appear older than the original faces produced based on the younger adult databases.

Primary analyses

The first analysis examined accuracy within training trials only. To determine possible fatigue effects, trials were organized into five sequential bins of 10 trials each, creating a five level within-subjects factor to add to the aging direction and feedback between-subjects factors. A Huynh-Feldt corrected within-subjects test found no effect of bins on accuracy, and thus no fatigue effects over time. Feedback yielded more accurate recognition than no feedback, $F(1, 144) = 7.79$, $p = .006$, $n^2_p = .051$. Direction produced no main effect (see Figure 10). Two interactions manifested. First, a bins x direction interaction was found, $F(4, 576) = 2.96$, $p = .02$, $n^2_p = .02$. Simple effects tests found the interaction was driven by greater accuracy for progression trials than regression trials, $F(1, 144) = 6.26$, $p = .01$, $n^2_p = .042$. Secondly, a bins x feedback interaction was found, $F(4, 576) = 3.60$, $p = .007$, $n^2_p = .024$. Simple effects tests found the interaction was driven by main effects of direction in Bin 4, $F(1, 144) = 18.18$, $p < .001$, $n^2_p = .112$, and Bin 5, $F(1, 144) = 6.30$, $p = .01$. In each case, progression training trials outperformed regression training trials.

The main analysis examining change from pretest to posttest used the same design as before. No effect of test item direction, training direction, or feedback was found. However, a test item direction x feedback interaction was found, $F(1, 144) = 5.97$, $p = .02$, $n^2_p = .04$ (see Figure 11). Simple effects tests revealed that the interaction was driven by the no feedback condition yielding greater posttest progression trial improvement than the feedback condition, $F(1, 144) = 8.21$, $p = .005$, $n^2_p = .054$.

Taken together, results from Experiment 2 show a clearer depiction of the effect of feedback, likely due to the removal of familiarity and image-level variability from stimuli. During training, feedback yielded greater accuracy overall than no feedback, respectively averaging 73% and 68% accuracy. Aging direction during training did not affect accuracy, which replicates the finding from Experiment 1 when one considers that there is no same age condition for EvoFIT faces. The interactions demonstrated that progression recognition generally improved over time, and importantly showed an improvement in later bins whereas no feedback accuracy remained constant. This is even more evident when one examines pretest accuracy, which for all groups averaged 65-67% percent and only improved over time with feedback. Unfortunately, this gain did not transfer to posttest accuracy, which yielded no effects except the 10% increase in posttest progression test trials in the no feedback condition.

Before speculating too much about the efficacy of feedback, two things are important to note that might be solved by making one modification to the design. First, test trials switch between progression and regression at random. Although no observed effects of direction appeared while direction was a between-subjects variable, in Experiments 1 and 2 progression accuracy was greater than regression accuracy at the within subjects level. Switching between progression tests and regression tests may tax working memory and the perceptual system, and progression may come easier simply because aging younger to older is what the visual system is more accustomed to. Second, all other investigations of improving face perception via feedback training have employed matching paradigms rather than the 2AFC paradigm employed in Experiments 1 and 2. A matching design would have no “direction” to confuse the perceptual system and would more accurately mimic the real-world task of identity matching. So, Experiment 3 stands as a direct replication of previous efforts at improving face recognition through training (e.g., White et al., 2014), albeit over an age gap. Matching paradigms, when properly constructed, also allow researchers to examine discriminability and response bias.

III. Experiment 3

A. Note

This experiment contains a serious error that renders its results’ interpretability questionable, and this note is included to ensure that individuals who may find this manuscript in a database understand this. Specifically, after data collection, initial analyses produced below-chance responses. This was due to Match trials and Mismatch trials’ feedback being switched such that correct trial decisions received negative feedback and incorrect decisions received positive feedback. Although recoding data for these analyses was simple, I cannot definitively assert why incorrect feedback reduced accuracy below chance (i.e., whether participants utilized new strategies or merely gave up).

B. Method

Participants

Fifty-four undergraduates across two universities participated in this study in exchange for credit toward a research participation requirement in their introductory psychology classes. Detailed information about the participants can be found in Table 1.

Materials, Design, and Procedure

Facial images were identical to those used in Experiment 2. The study employed a 2 (Feedback vs. Not) x 2 (Trial Type: Match vs Mismatch) mixed design, with trial type as a within-subjects factor. Like Experiments 1 and 2, participants took a 20-trial pretest and posttest, which were counterbalanced, with an intervening 50-trial training regimen. The study employed a matching paradigm rather than a two-alternative forced choice delayed recognition paradigm. To this end, each randomly presented trial consisted of a fixation point followed by two faces side-by-side on the screen. Participants were instructed to press the "Q" key on the keyboard if the faces belonged to the same identity and the "P" key if they were different identities. Half of the trials contained matched identities, and half contained mismatched identities. After sessions concluded, participants were debriefed and dismissed.

C. Results and Discussion

The main dependent variables of interest are the same as those from Experiment 1. Again, no effects of counterbalancing group on overall accuracy were found, with accuracy averaging 62% and 64%. Collection site also had no effect, averaging 62% and 65%. Participant gender was unrelated to overall accuracy. A negative relationship of $r = -.33$ between participant age and overall accuracy did manifest, $p = .02$. However, five participants failed to enter an age (resulting in an age of 0) and several mistakenly entered single digits of 1 or 2 rather than their complete age. After removing these cases, the significant relationship disappeared.

Primary Analyses

The first analysis examined match and mismatch accuracy rates during training. Match trials were more accurate than mismatch trials, $F(1, 52) = 64.69$, $p < .001$, $\eta^2_p = .554$. No feedback also outperformed feedback, $F(1, 52) = 4.44$, $p = .04$, $\eta^2_p = .079$. No interaction manifested (see Figure 12). Because match and mismatch trials were presented at random throughout training, splitting training into bins is not appropriate to estimate possible improvement over time. So, a simple bivariate correlation was carried out, showing that trial order and accuracy produced a marginal negative relationship, $r = -.25$, $p = .08$ (see Figure 13).

The primary analysis of interest was again whether posttest trials saw improvement compared to pretest trials. No effect of trial type manifested here; however, no feedback yielded greater improvement than feedback, $F(1, 52) = 14.25$, $p < .001$, $\eta^2_p = .215$. No interaction was found (see Figure 14).

Response Bias

One advantage to matching paradigms over 2AFC is that response bias can be calculated. A simple response bias measure, Q, was calculated for each participant for pretest, posttest, and training responses and used as a dependent measure for the following analyses. Q determines what proportion of total errors from match trials (i.e., misses) and mismatch trials (i.e., false alarms) are matching errors:

$$Q = \frac{misses}{misses + false\ alarms} \tag{1}$$

Q is also simple to interpret, where $Q \approx .50$ signifies roughly equitable yes and no responses, $Q > .50$ signifies conservative bias, and $Q < .50$ signifies liberal bias.

No effect of feedback was found on response bias during pretests or posttests, with Q values generally indicating equitable responding during pretests and posttests in both feedback and on feedback regimens. During training, no feedback response bias (*M* = .27, *SD* = .19) was more liberal than feedback (*M* = .36, *SD* = .12), $F(1, 52) = 3.88$, $p = .05$, $\eta^2_p = .069$. Response bias during training (*M* = .63, *SD* = .13) was generally liberal when compared to the equitable value of .50, $t(53) = 7.39$, $p < .001$.

With Experiment 3, evidence for a feedback effect on face matching manifested, albeit in an unintentional direction. Systematically incorrect feedback during training reduced face matching accuracy on post-tests compared to the no feedback control condition. During training itself, feedback averaged approximately 57% accuracy whereas no feedback averaged 63%. Feedback yielded a posttest decrement of approximately 10% whereas no feedback yielded a posttest gain of about 7%. Moreover, these can genuinely be attributed to an increase in discriminability rather than response bias, indicating a strategic readjustment of weighting perceptual cues among those who received feedback compared to those who did not. Further implications regarding feedback will be discussed in greater detail in the discussion.

V. General Discussion

Overall, the studies presented here revealed that face recognition and matching improved after training, particularly among artificial faces that could not be familiar to participants. Feedback more greatly affected matching than recognition. Because overall patterns have been reiterated in previous sections, this general discussion will focus on how theories of recognition memory, perceptual learning, and feedback interventions apply to the current set of studies. After this, potential future directions will be outlined.

A. Face Memory and Perceptual Learning

Perceptual learning regimens work on the general principle that controlled processing of information demands more attention and elaboration than automatic processing, and that monitored practice leads skills requiring the former to rely on the latter (e.g., Schneider & Schiffrin, 1977). Expertise in a domain might even be conceptualized as the transfer of controlled processes for a given task to automatic processes. Face recognition is something for which human beings have great expertise (Wallis, 2013), primarily relying on automatic processes in most day to day circumstances. Even cases where recognition rates decline (such as during cross-race recognition), they still average above chance. In the current case, even age gap recognitions averaged above chance despite major morphological changes that come with natural aging. It is worth pointing out here that the matching task averaged near or below chance accuracy. This could be explained by the brief encoding time afforded during the recognition tasks, followed by two choices that participants know include the correct choice. With only two seconds to encode faces, participants relied on their life-long developed automatic face processing and were then better equipped to respond at test. Similarly, match trials tended to produce more accurate responses than mismatch trials because match cues are easier and more automatically detected than mismatch cues across several perceptual domains.

Related to automaticity, another perspective on the current results comes in light of recognition memory theory. In many models and theories of recognition memory (Jacoby, 1991; McClellend & Chappel, 1998; Yonelinas, 1994), familiarity is considered automatic whereas recollection requires more control and working memory. In the recognition task utilized in Experiments 1 and 2, participants already had a chance to see a version of the face they would be tested on. Therefore, one of the two test faces by default would be more familiar than the other, even if both portray a different age than the study face. McClellend and Chappel (1998) point out that this process might also rely on differentiation. For a single item to be recognized, not only must its similarity to a studied item be detected but it must also be differentiated from unstudied items. Effectively, the 2AFC tasks used in Experiments 1 and 2 allowed participants to study faces twice, and in cases of age gaps, they are narrowed to only the most basic feature cues that remain common throughout a lifetime. This is something that participants already perform well at, but with practice general improvements developed. Although matching does not allow for familiarity and differentiation the way that recognition tasks do, in principal the presence of match and mismatch cues allows for participants to take similarities and differences into account in real time without the constraints of encoding and short-term decay.

More recently, visual perceptual learning has been found to work best when learners get to begin or practice with easy trials or examples and only then proceed to more difficult trials (Ahissar & Hochstein, 1997). This "reverse hierarchy" works by drawing attention to task-relevant features more obvious in the easy trials. The learner can then transfer these strategies more easily difficult trials compared to beginning with difficult trials. This particular method of administering perceptual learning was not employed in the current experiments, although the data collected does provide normative information for which trials are generally more difficult and which are easier, as is apparent in Appendix B. Future implementation of the current training paradigm can focus on such a transfer to determining if training sequenced from easy to difficult produces robust posttest improvement over pretest.

B. The Feedback Paradox

The current manuscript presents multiple instances where feedback interventions (FIs) effectively improve face recognition and matching and several instances where they do not. Although this mixture muddles interpretability of the three experiments presented, it nonetheless captures the variability of success (or lack thereof) for feedback training found throughout the literature studying such interventions (Kluger & DeNisi, 1996). FIs have been at the center of theoretical and applied research for almost as long as psychology has endured scientific investigation. They have been tested to improve performance in educational settings (Schloss, Wisniewski, & Cartwright, 1988), memory (Titus, 1973), problem-solving (Crafts & Gilbert, 1935), therapy (Baechle & Lian, 1990), and industry (Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1988). Such is the breadth of its use in empirical investigations that a Google Scholar search of the exact phrase “feedback intervention” including the word “performance” yields over 5,300 results at the time of this writing. Although the current paper is only one of a few tests of its use in face recognition paradigms, consideration of its use in other domains may explain why it did and did not work here.

Thorndike’s (1927) law of effect and similar behavioristic accounts underlie the hypothesis that FIs improve performance. In other words, following an action with a desirable outcome increases the probability that that action will be performed again, whereas following an action with and undesirable outcome reduces its likelihood of repetition. This account is merely descriptive and does not elucidate process or sub-mechanisms. Kluger & DeNisi (1996) provide a formal definition of FIs as “actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance” (p. 255). FIs in general operate on attention and motivation, thereby increasing effort by narrowing attention to task-related actions that result in success or failure. Importantly, the authors divide feedback into two major subcategories: feedback that supplies knowledge of results (KR) and feedback that overtly explains how to perform a task better. Regardless of type, the process model they propose results in multiple, separate possible outcomes. Positive feedback motivates individuals to increase effort when they are presented with a standard to reach. Negative feedback can also increase effort. However, if the feedback comes with no explanation or does not result in improved performance, it can reduce effort and shift attention away from the task and toward the self in a way similar to learned helplessness. Simply put: Feedback can help individuals who are doing well but sub-optimally, but feedback can discourage individuals who are doing poorly.

Aside from modelling the mechanisms driving FIs, Kluger & DeNisi (1996) conducted a meta-analysis on the major moderators examined in the literature up to that time. Relevant to the current experiments, mere correct-incorrect feedback like that used in the current studies negatively affects performance, (d = -.13), attainment level (analogous to the current study's cumulative proportion correct presented along with feedback) does not affect performance, computerized feedback has no effect, task time constraints (such as the short encoding time for the recognition studies) reduce performance (d = -.11), memory tasks in general do not benefit from feedback, and lab-based studies such as those currently presented are less likely to see performance improvement than field studies (d = -.17). For practical reasons, the current experiments used KR feedback because it is impossible to know given the current data why individuals made errors in any given trial. Specifically, participants may have made erroneous responses based on any given aspect of familiarity, expression, pose, lighting, or external features that could have triggered a sense of perceptual similarity. Failure to observe a consistent feedback effect among the current experiments may be rooted in correct-incorrect trial-by-trial administration, reducing motivation among those who are already performing near ceiling and also reducing motivation among those who are doing poorly yet not receiving substantive feedback to change strategies. Switching from a 2AFC paradigm to a matching paradigm in Experiment 3 eliminated the memory component as well as the time constraint on encoding, thus yielding a clear feedback effect during and after training otherwise inconsistently observed in Experiments 1 and 2. For the purposes of the current manuscript, lab-based administration was unavoidable. Moreover, computer-based administration is more desirable than field administration given that one end goal of the current research is to develop a portable system designed to augment training of law enforcement and security officials. These and other future directions will be discussed in more detail below.

C. Implications for Law Enforcement

Facial appearance is one of many biometric cues used by law enforcement to track and identify suspects. Other such information includes fingerprints, palmprints, iris patterns, voice, DNA, and the technology developed to analyze and identify these elements (National Science and Technology Council, 2011). Expert face analysts are more likely to accurately identify unfamiliar faces in optimal, same-age conditions where faces are roughly the same appearance in multiple images (Wilkinson & Evans, 2009). Given that same-age recognitions were observed here in Experiment 1 to be proximal to recognition rates found elsewhere (Burton et al., 2010), it would be reasonable to predict that experts would recognition faces after an aging lapse at a similar rate of increase. Of course, high error rates would still manifest resulting in potentially large numbers of fugitives eluding apprehension.

The United States' Federal Bureau of Investigations includes general education about facial morphology and physiology and even includes identity match training as a standard for their agents (Bruegge, 2011). However, this training typically involves comparing face images of varying qualities sourced from various appearance changes and capture methods (e.g., close-circuit television, ID photos, confiscated personal photos, etc.). Facial age, however, is not explicitly addressed in these regimens. Given the rates of age gap recognition rates observed in Experiments 1 and 2, and the comparatively lower rates of age gap face matching in Experiment 3, this would be cause for alarm to law enforcement agencies charged with tracking down long-term missing persons and fugitives. Of course, it is likely that investigators in real cases have access to multiple images of targets. Having multiple images of an individual to study and have available for reference would generate more recognition cues than single images. As described above, these cues might increase recognition rates even after an aging lapse compared to those generated by single images.

D. Future Directions

Overall, the three studies show promise for future investigations of face recognition and matching training. One shortcoming of the current study lies in the fact that face stimuli used in Experiment 1 were not only uncontrolled for pose and expression, but that their identities' celebrity status made it difficult to be certain that participants were unfamiliar with them. This is more apparent given that even though we prescreened our participants' familiarity, they nonetheless found some faces familiar. There are two ways to combat this problem in future investigations. First would require finding a naïve population who would be unfamiliar with the facial identities. This population could be taken from another region on Earth where American and British celebrities are less well-known or sourced in the future when our college-aged participants would be less likely to be familiar with 20th Century celebrities. The other way would require developing a large-scale database of face identities including volunteers whose face images are collected and taken over time. This too would require many years and decades, but such a database would be advantageous over artificial faces for improved ecological validity and over celebrity faces for control over pose, lighting, and expression.

Another avenue for future research relates to improving recognition of missing children. Almost half of the missing persons reports in 2013 involved minors (NCIC, 2013), and over 40% of missing children's cases involve children missing for more than five years (Lampinen et al., 2009). Given the tendency for age-progressions of missing children to yield recognitions equivalent to or less accurately than outdated images, attempts to improve recognition and matching of adult images based on childhood images would be a worthwhile investigation. Assembling a database of such photographs would be simple because portrait-quality photographs of children are taken by public schools each year.

Future research could also manipulate feedback in different ways than the current study. For example, as Kluger & DeNisi (1996) uncovered in their meta-analysis, mere KR feedback has a weak negative effect on performance, and is less desirable than feedback implementing how performance failed to meet a standard. As stated above, it is impossible to know prescisely what kind of error a person makes when making the wrong identity judgment. However, specialized instructions could be provided before training or alongside feedback recommending new strategies (e.g., "make your judgment based on the upper face region, which remains most constant over time"). The presentation of cumulative feedback could also be made more useful by providing a clear standard for participants receiving feedback to attain (e.g., "Your total accuracy is 72%! Keep going until you reach 90%.").

Finally, basic and theoretical investigations could be designed to determine the mechanisms by which the human visual system is able to identify unfamiliar individuals based on other-age study photos. Such research could be completed after assembling adequate standardized images of faces at various ages. These investigations would augment current face recognition theories or aid in the creation of new ones to guide future research.

References

Abdi, H., Valentin, D., & Edelman, B. G. (1998). Eigenfeatures as intermediate-level representations: The case for PCA models. *Behavioral and Brain Sciences, 21*, 17-18.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401-406.

Albert, A. M., Ricanek, K., & Patterson, E. (2007). A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, *172*, 1-9.

Baechle, C. L., & Lian, M. G. J. (1990). The effects of direct feedback and practice on metaphor performance in children with learning disabilities. *Journal of learning disabilities, 23*, 451-456.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77,* 305-327.

Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition*, *19*, 221-228.

Bruegge, R. V. (2011). FBI facial comparison and identification training. *Paper presented at the 2011 Biometric Consortium Conference, Tampa, FL.*

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*, 286-291.

Crafts, L. W., & Gilbert, R. W. (1935). The effect of knowledge of results on maze learning and retention. *Journal of Educational Psychology, 26*(3), 177-187.

Dakin, S. C., & Omigie, D. (2009). Psychophysical evidence for a non-linear representation of facial identity. *Vision research*, *49*, 2285-2296.

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British journal of psychology, 106,* 433-445.

Enlow, D. H., & Hans, M. G. (1996). *Essentials of facial growth*. Philadelphia, PA: Saunders.

Erickson, W. B., Lampinen, J. M., Frowd, C. D., & Mahoney, G. (2013). [Use of morphed age-progressions in simulated missing persons cases]. Unpublished raw data.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception?. *Psychological review*, *105,* 482-498.

Frowd, C. D., Bruce, V., McIntyre, A. H., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P. J. (2006). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, *1*, 42-51.

Frowd, C. D., Hancock, P. J., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on applied perception (TAP)*, *1*, 19-39.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision research*, *37*, 1673-1682.

Gillgrass, T.J., & Welbury, R. (2012). Craniofacial growth and development. In R. Welbury, M. S. Duggal, & M. T. Hosey (Eds.), *Paediatric Dentristry* (1-17). Oxford: Oxford University Press.

Goodnough, A. (2011, June 27). Crime boss in hiding found time for travels. *The New York Times.* Retrieved from http://www.nytimes.com/2011/06/28/us/28bulger.html

Guyuron, B., Rowe, D. J., Weinfeld, A. B., Eshraghi, Y., Fathi, A., & Iamphongsai, S. (2009). Factors contributing to the facial aging of identical twins. *Plastic and reconstructive surgery*, *123*, 1321-1331.

Hancock, P. J., & Little, A. C. (2011). Adaptation may cause some of the face caricature effect. *Perception*, *40*(3), 317-322.

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology*, *46*, 445-448.

Hole, G., & George, P. (2011). Evidence for holistic processing of facial age. *Visual Cognition*, *19*, 585-615.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Vol. 1, No. 2, p. 3). Technical Report 07-49, University of Massachusetts, Amherst.

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision research*, *49*, 2273-2284.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313-323.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*, 254-284.

Lampinen, J., Arnal, J. D., Adams, J., Courtney, K., & Hicks, J. L. (2012). Forensic age progression and the search for missing children. *Psychology, Crime & Law*, *18*, 405-415.

Lampinen, J.M., Arnal, J.D., Courtney, K., & Adams, J. (2009). Archival study of cases listed on the website of the National Center for Missing and Exploited Children. *Annual Meeting of the Midwestern Psychological Association. Chicago IL.*

Lampinen, J. M., Erickson, W. B., Frowd, C. D. & Mahoney, G. (2015). Mighty morphin' age progressions. *Psychology, Crime, & Law, 21* 952-967.

Lampinen, J. M., Miller, J. T., & Dehon, H. (2012). Depicting the missing: Prospective and retrospective person memory for age progressed images. *Applied Cognitive Psychology, 26*(2), 167-173.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*, 865-876.

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.

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.

Moynihan Ramsey, K., Marcheva, B., Kohsaka, A., & Bass, J. (2007). The clockwork of metabolism. *Annual Review of Nutrition*, *27*, 219-240.

National Center for Missing and Exploited Children (2016). *NCMEC by the numbers.* Retrieved from http://www.missingkids.org/KeyFacts

National Crime Information Center (2013). *NCIC Missing Person and Unidentified Person Statistics for 2013.* Retrieved from https://www.fbi.gov/about-us/cjis/ncic/ncic-missing-person-and-unidentified-person-statistics-for-2013

National Science and Technology Council (2011). *The national biometrics challenge.* Retrieved from http://www.biometrics.gov/Documents/BiometricsChallenge2011_protected.pdf

O'Toole, A. J., Phillips, P. J., Jiang, F., Ayyad, J., Penard, N., & Abdi, H. (2007). Face recognition algorithms surpass humans matching faces over changes in illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*, 1642-1646.

Pittenger, J. B., & Shaw, R. E. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(4), 374-382.

Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, *73*, 337-358.

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception*, *19*, 773-794.

Ricanek, K., & Tesafaye, T. (2006). MORPH: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on* (pp. 341-345). IEEE.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*, 1281-1302.

Sadler, L. L. (1986). Scientific art and the milk carton kids. *Paper presented at 7th Annual Guild of Scientific Illustrators, Washington, DC*.

Schloss, P. J., Wisniewski, L. A., & Cartwright, G. P. (1988). The differential effect of learner control and feedback in college students' performance on CAI modules. *Journal of Educational Computing Research*, *4*, 141-150.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, *84*, 1-66.

Seamon, J. G. (1982). Dynamic facial recognition: Examination of a natural phenomenon. *American Journal of Psychology, 95*, 363-381.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Pearson

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*, 225-245.

Taylor, K. T. (2000). *Forensic art and illustration.* Boca Raton, FL: CRC Press.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*, 212-222.

Titus, T. G. (1973). Continuous feedback in recognition memory. *Perceptual and Motor Skills, 37*, 771-776.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience, 3*, 71-86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*, 161-204.

Wallis, G. (2013). Toward a unified model of face and object recognition in the human visual system. *Frontiers in psychology*, *4*, 1-25.

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic bulletin & review*, *21*, 100-106.

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images?. *Science & Justice*, *49*, 191-196.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.

Table 1. Characteristics of samples.

Characteristic	**Experiment 1**	**Experiment 2**	**Experiment 3**
Participants	237	158	54
Collection Site			
University of Arkansas	131	65	30
Arkansas State University	106	83	24
Mean Age (Standard Deviation)	19.15 (8.01)	18.86 (5.71)	18.53 (4.07)
Sex			
Female	179	113	35
Male	58	35	19
Race			
White/Caucasian	164	97	36
African American/Black	37	17	5
Hispanic/Latino	16	12	4
Asian/Pacific Islander	13	15	7
Native American	3	1	1
Other	4	3	1

Table 2. Distribution data for each age range and direction for pairings in Experiment 1, collapsed across training and test trials.

Direction/Range	**N**	***M***	***SD***	**Min**	**Max**	**Skewness**	**Kurtosis**
Gap (Progression)	90	62.77%	9.80	39%	84%	0.11	-0.30
Gap (Regression)	90	65.86%	10.47	35%	87%	-0.45	0.24
Same Age (Older)	50	84.74%	10.55	47%	100%	-1.41	3.37
Same Age (Younger)	50	75.98%	14.91	30%	98%	-1.08	1.66

Table 3. Self-reported number of faces participants recognized (if participants recognized more than nine, they were told to input nine and indicate the true number in the free recall naming section).

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
How many faces did you recognize? (0-9)	70	39	38	36	19	21	10	1	0	3

Table 4. Familiarity survey responses.

Familiarity Probe	**Responses**
List the names of any persons that you think you recognized below. If you did not recognize anyone, type "None".	
None*	96
Indicated familiarity, but no names	21
Provided at least one name	120
How many classic films (1930s-1960s) have you seen?	
None	105
1 to 5	122
6 to 15	9
15+	1
How many classic TV shows (1960s - 1980s) have you seen?	
None	85
1 to 5	141
6 to 15	11
15+	0
How many classic rock musicians' appearances (1970s-1990s) are you familiar with?	
None	111
1 to 5	104
6 to 15	20
15+	2

**More participants responded "none" to this question than responded "0" for the recognition question in Table 2. This discrepancy is likely due to some participants entering "none" rather than choose not to indicate general familiarity.*

Figure 1. Graphical recreation of Seamon's (1982, Experiment 5) results demonstrating robust ability to match adult faces to same-identity childhood images.

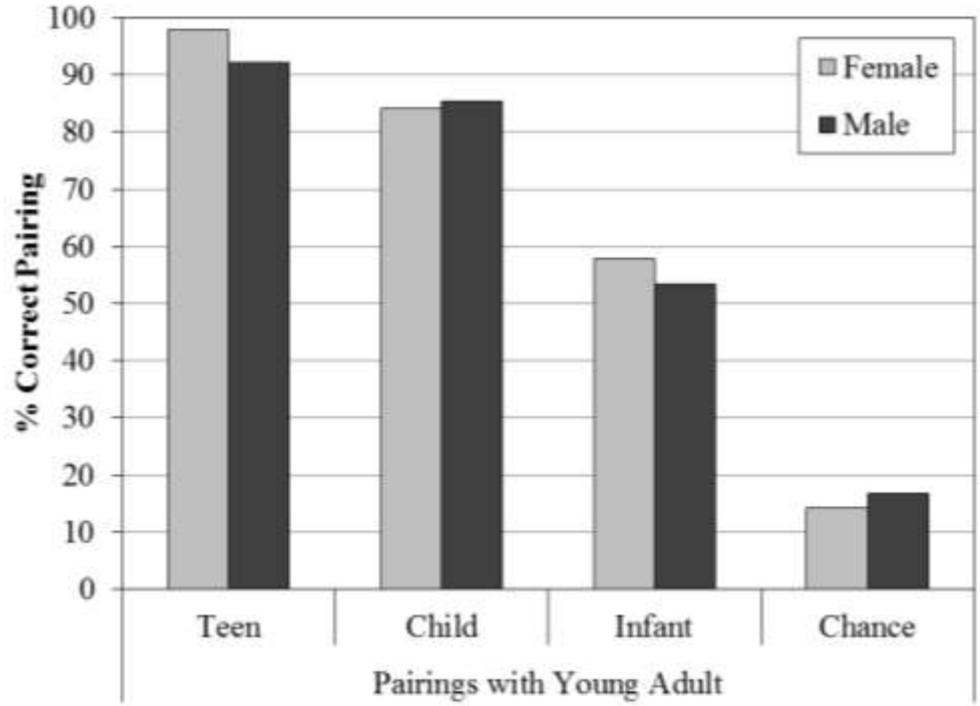


Figure 2. Data from Erickson, Lampinen, Frowd, & Mahoney (2013) showing that childhood study images produce marginally greater recognition of adult images than various age progression techniques. Difference scores calculated from subtracting proportion of false alarms from proportion of hits. Error bars represent 95% confidence intervals.

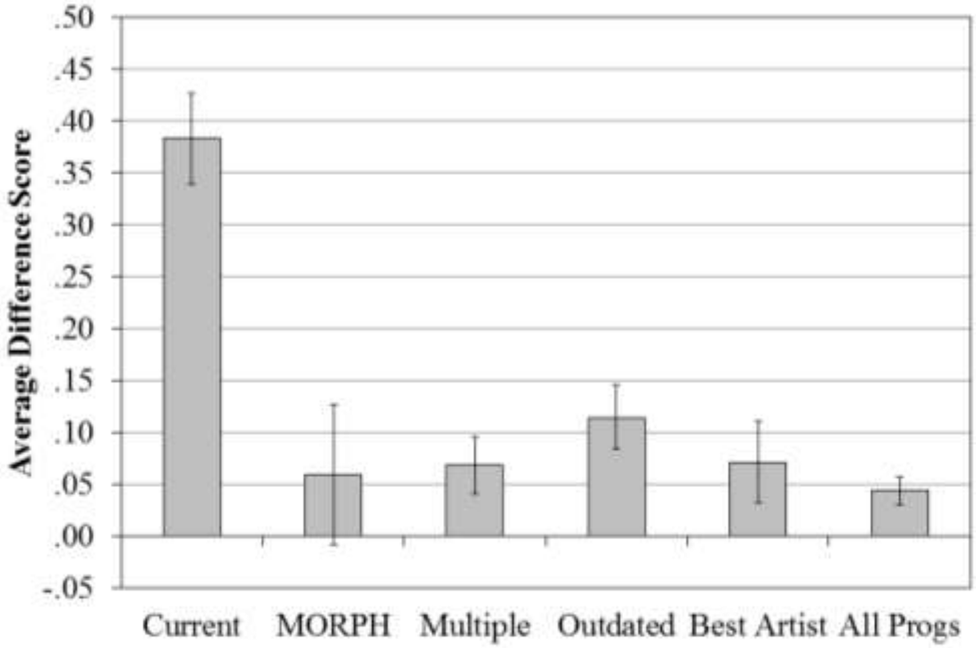


Figure 3. Example trial sequence from Experiment 1 displaying a progression trial beginning with study, then proceeding to mask, and ending with test.

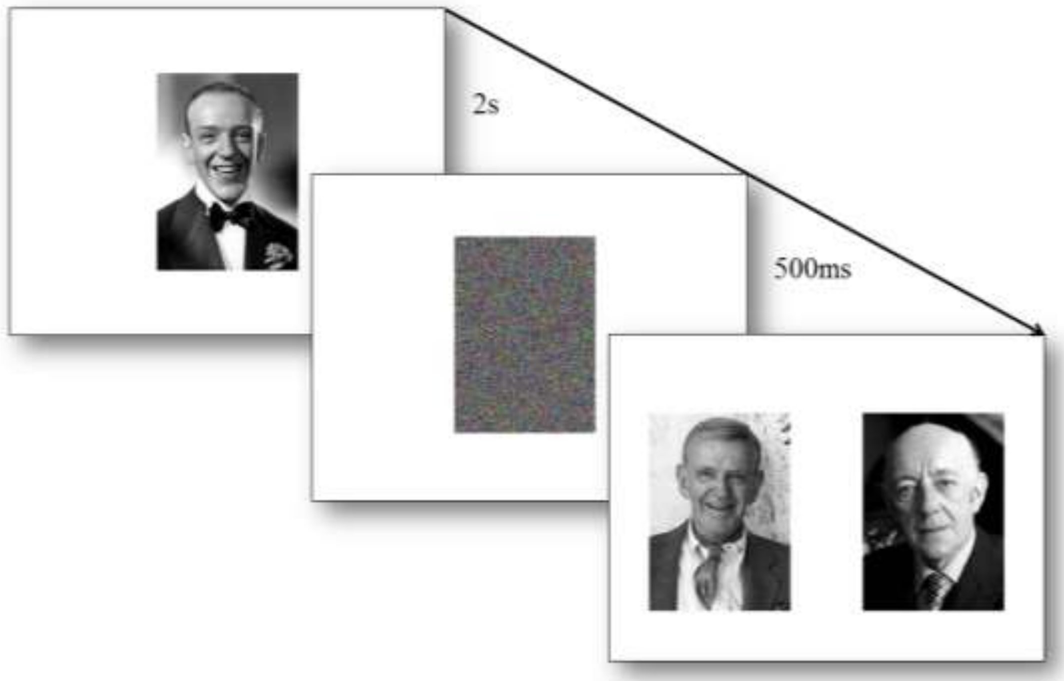


Figure 4. Schematic representation of between-subjects factors in experimental sessions in Experiment 1.

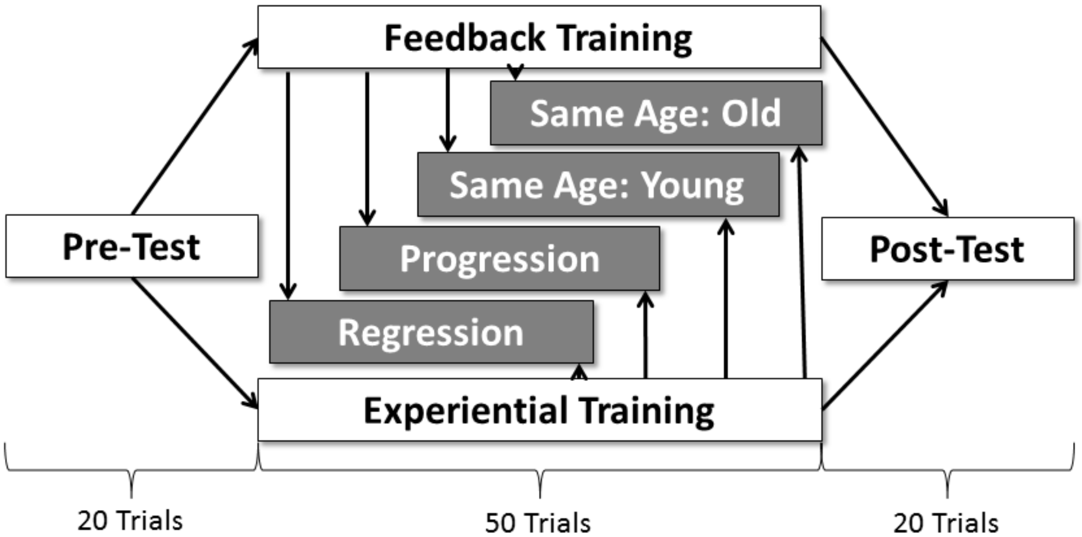


Figure 5. Training accuracy rates incorporating data from the entire sample in Experiment 1. Error bars represent 95% confidence intervals.

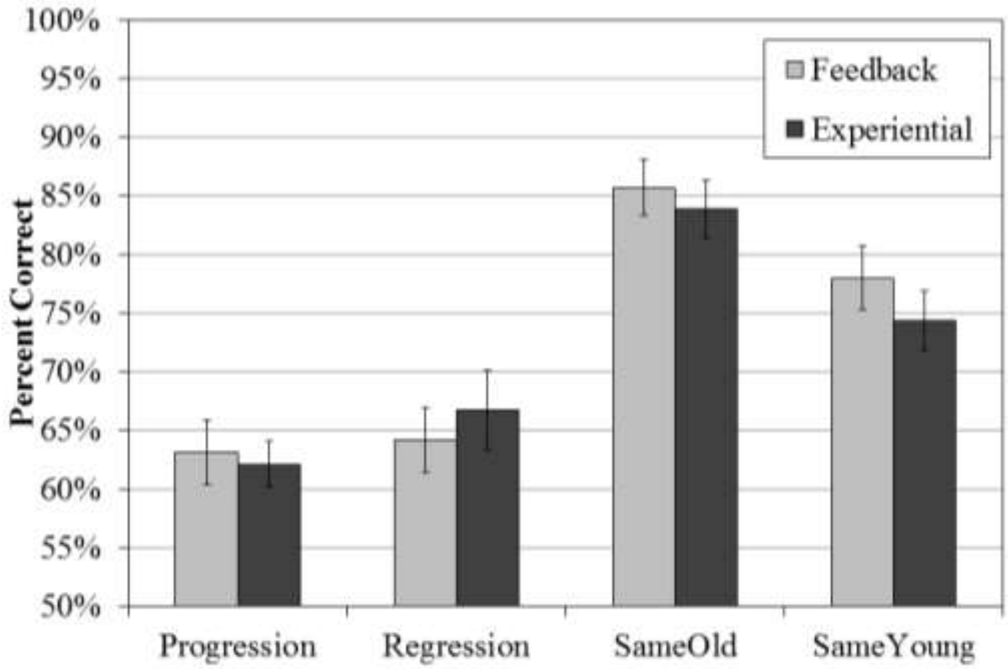


Figure 6. Difference score data incorporating data from the entire sample in Experiment 1. Error bars represent 95% confidence intervals.

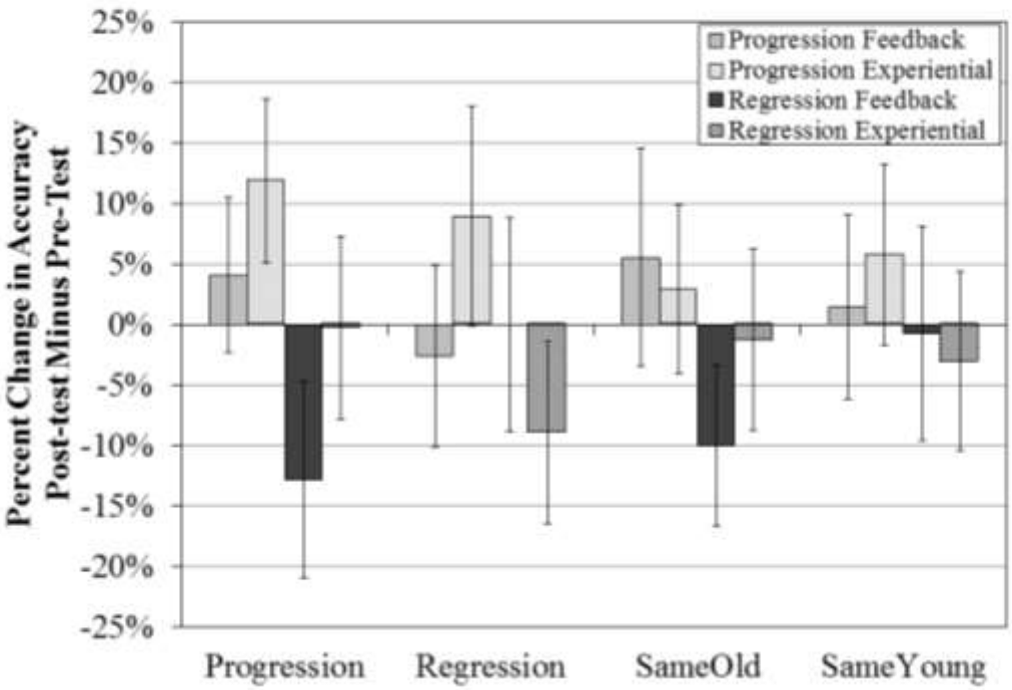


Figure 7. Training accuracy across feedback and age conditions for Caucasian participants in Experiment 1 only. Data are collapsed across trial bins for clarity. Error bars represent 95% confidence intervals.

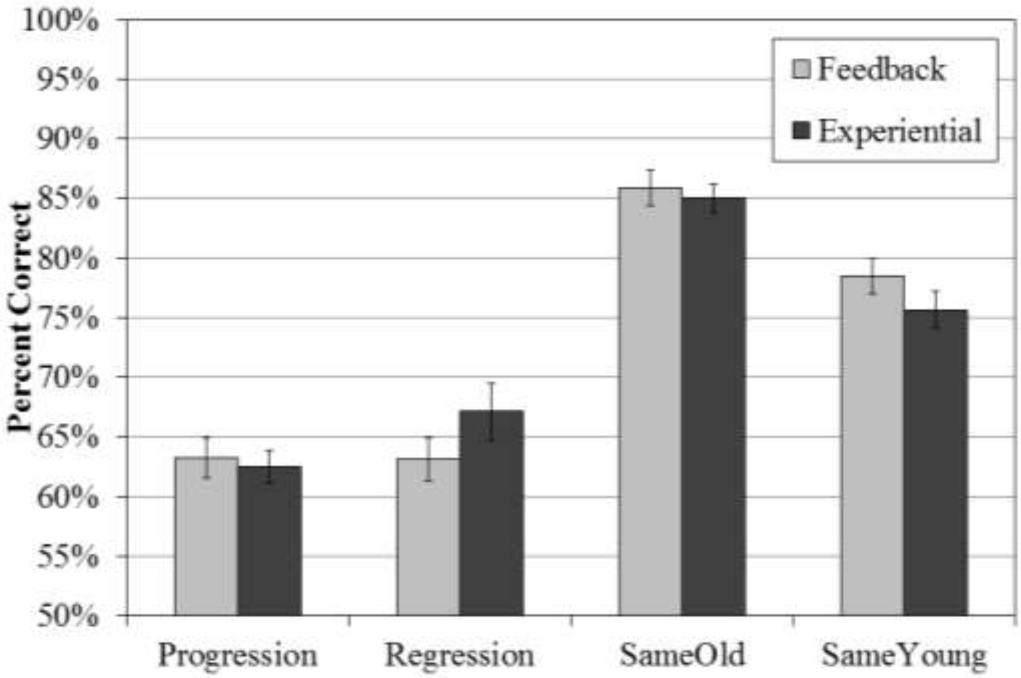


Figure 8. Percentage differences between posttest and pretest in the analysis including only Caucasian participants from Experiment 1. Error bars represent 95% confidence intervals.

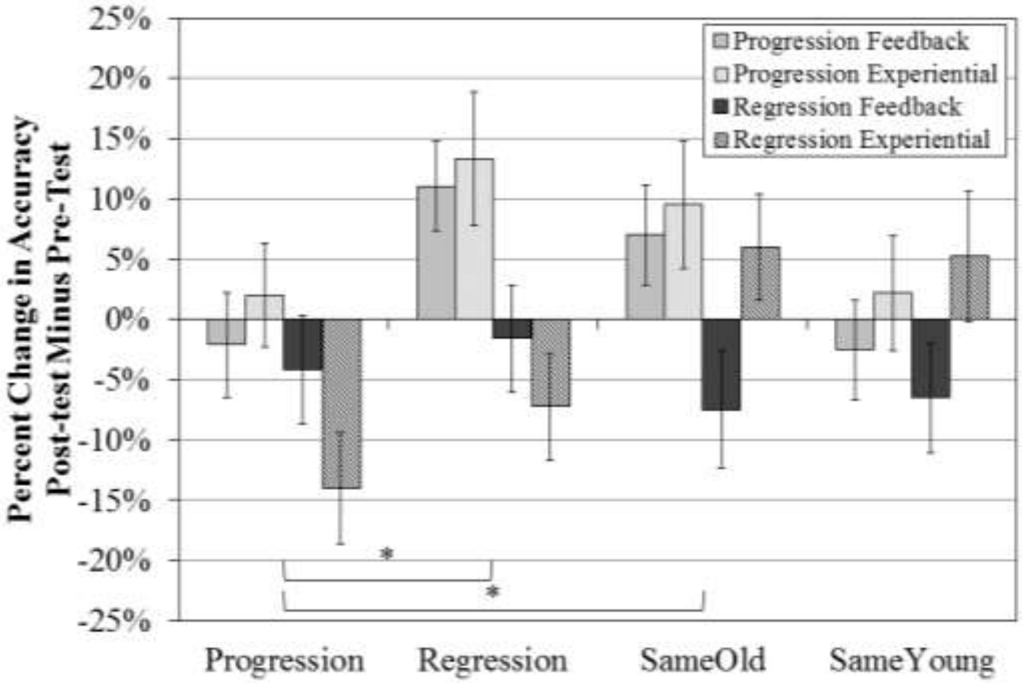


Figure 9. Examples of two stimulus EvoFIT identities used as stimuli in Experiments 2 and 3: (a) young female, (b) older female, (c) young male, and (d) older male.

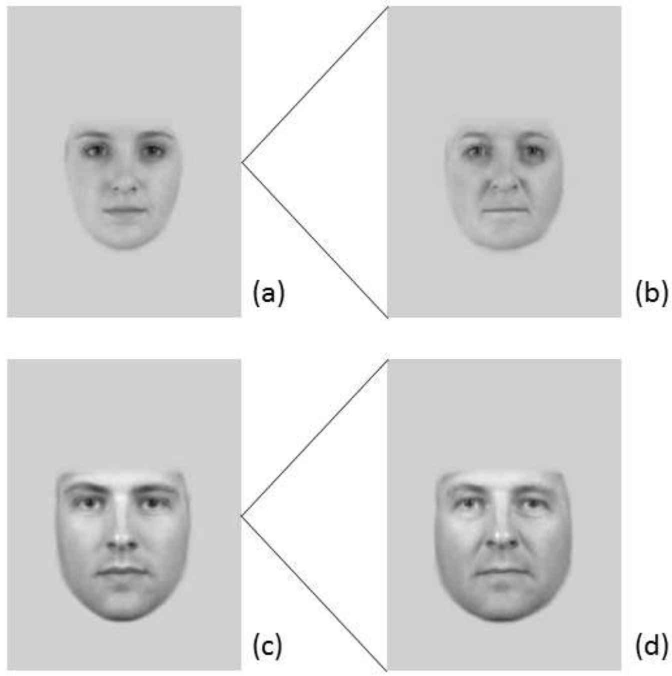


Figure 10. Accuracy rates during training in Experiment 2. Error bars represent 95% confidence intervals.

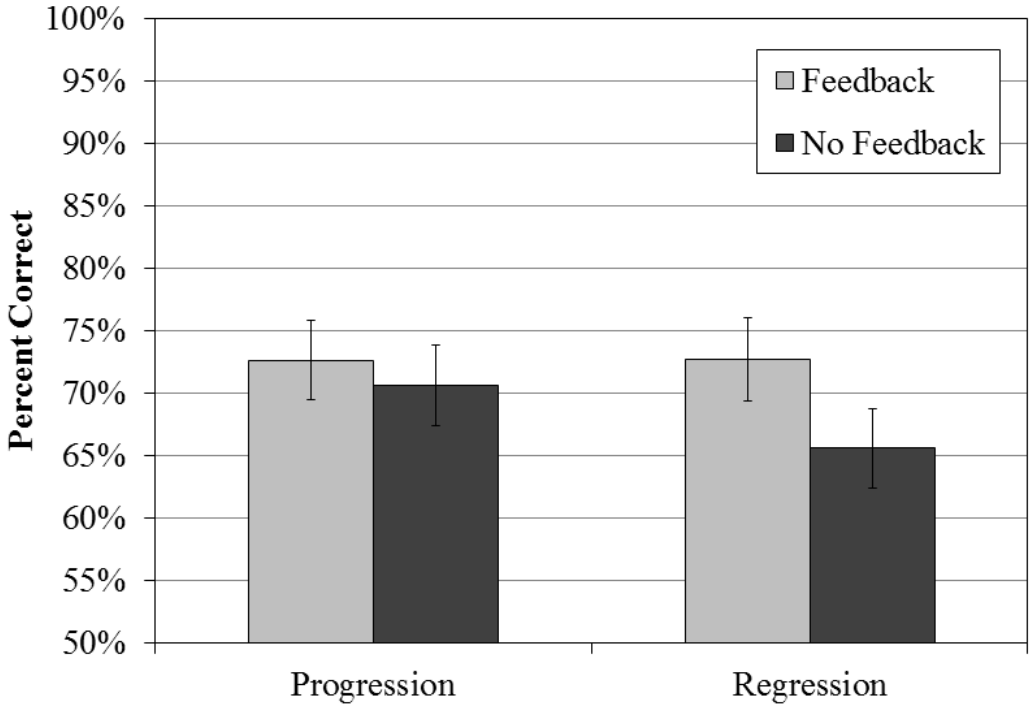


Figure 11. Percentage of change between pretest and posttest in Experiment 2. Error bars represent 95% confidence intervals.

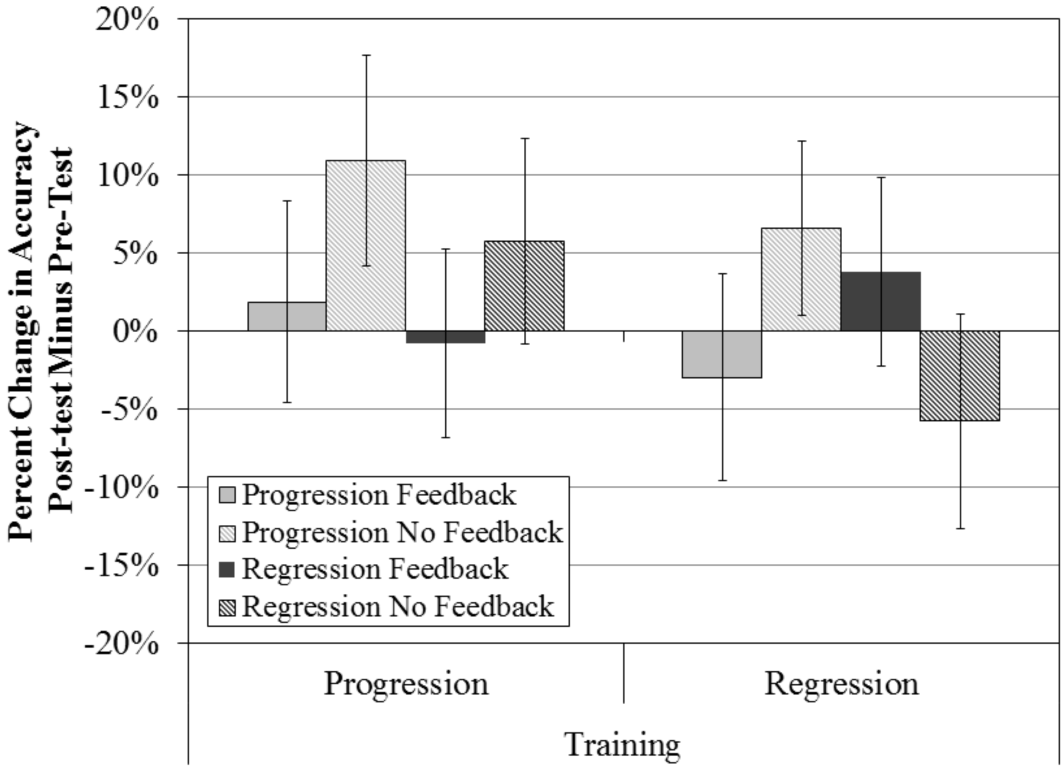


Figure 12. Mean correct matching judgments during training in Experiment 3. Error bars represent 95% confidence intervals.

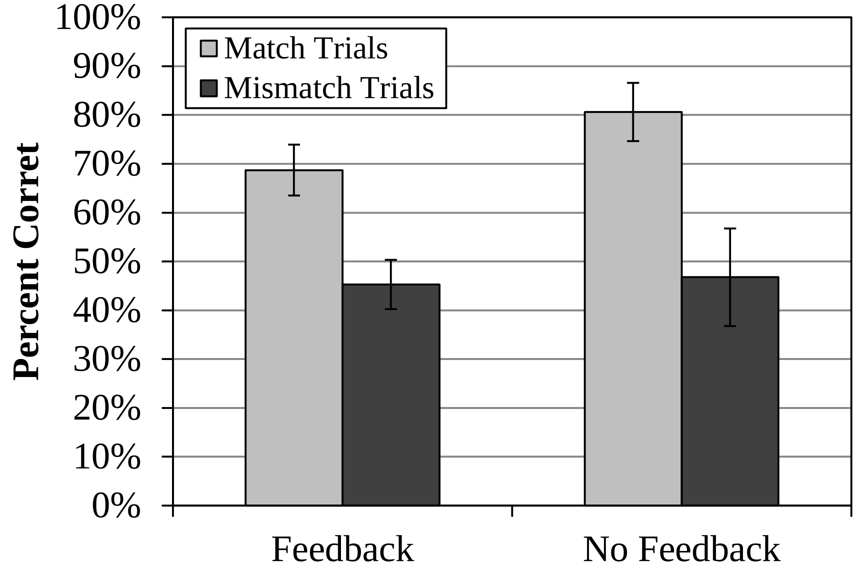


Figure 13. Scatterplot with trendline charting trial number against average accuracy per trial in Experiment 3.

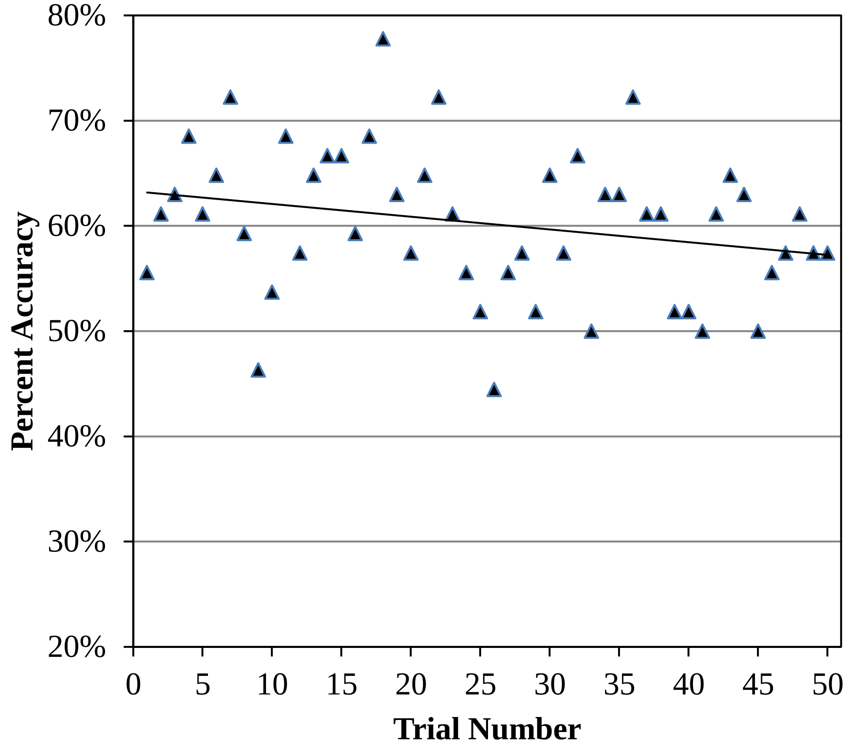
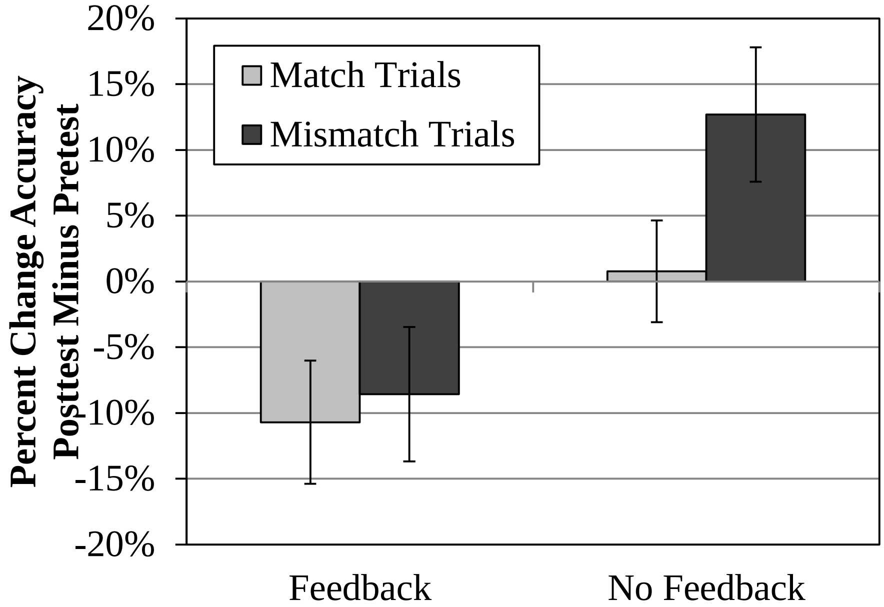


Figure 14. Percent change in accuracy from pretest to posttest in Experiment 3. Error Bars represent 95% confidence intervals.



Appendix A

Prescreening questionnaire for Experiment 1

1. How familiar are you with the faces of movie actors and actresses from the 1930s through the 1960s?
    a. extremely unfamiliar
    b. moderately unfamiliar
    c. somewhat unfamiliar
    d. somewhat familiar
    e. moderately familiar
    f. extremely familiar

2. How familiar are you with the faces of television actors and actresses from the 1960s through the 1980s?
    a. extremely unfamiliar
    b. moderately unfamiliar
    c. somewhat unfamiliar
    d. somewhat familiar
    e. moderately familiar
    f. extremely familiar

3. How familiar are you with the faces of classic rock musicians from the 1960s through the 1990s?
    a. extremely unfamiliar
    b. moderately unfamiliar
    c. somewhat unfamiliar
    d. somewhat familiar
    e. moderately familiar
    f. extremely familiar

Appendix B

Pairing identities and recognition rates for each face identity in Experiment 1

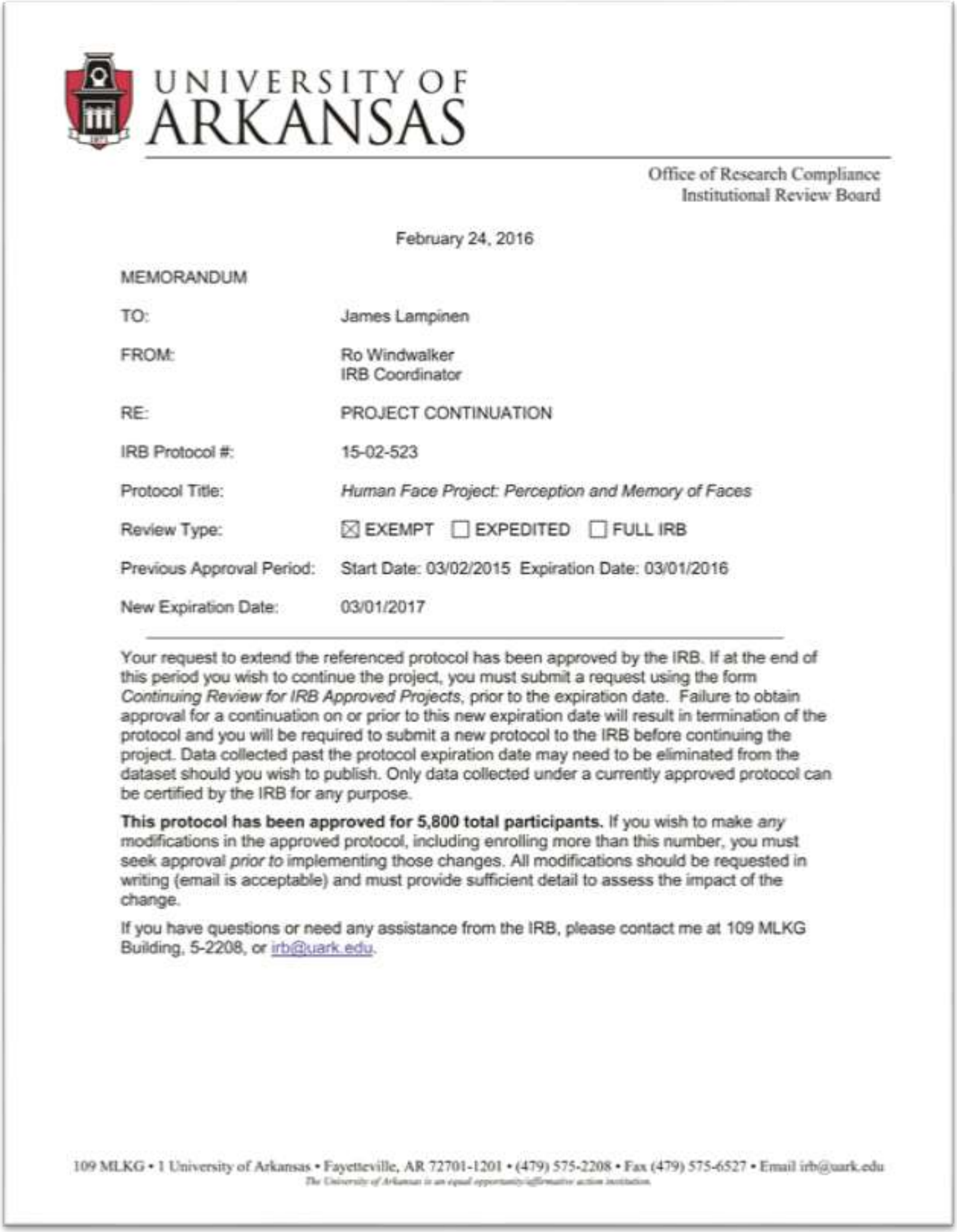
Pairing	Identity A	Identity B	Progression	Regression	Older	Younger
Training						
M01	George Kennedy	James Mason	75%	78%	90%	98%
M02	Anthony Franciosa	Anthony Quinn	52%	55%	84%	63%
M03	Eddie Vedder	Chris Cornell	63%	67%	84%	52%
M04	Graham Nash	Peter Tork	61%	65%	86%	48%
M05	Steven Stills	David Gilmour	59%	75%	91%	80%
M06	Dennis Wilson	Brian Wilson	66%	75%	53%	72%
M07	Noel Gallagher	Liam Gallagher	63%	65%	86%	75%
M08	Paul Simon	Peter Criss	55%	67%	97%	68%
M09	Roger Waters	John Fogerty	64%	56%	88%	80%
M10	Ritchie Blackmore	Gene Simmons	59%	75%	81%	65%
M11	Vince Neil	David Lee Roth	64%	76%	83%	72%
M12	Peter Frampton	John Paul Jones	58%	53%	97%	30%
M13	Robert Plant	Sammy Hagar	58%	67%	100%	97%
M14	Tom Petty	Tommy Shaw	78%	49%	84%	77%
M15	Steve Perry	Ozzy Osbourne	41%	69%	47%	58%
M16	Kris Novoselic	Nick Mason	73%	80%	86%	97%
M17	Bon Jovi	Axel Rose	47%	65%	74%	80%
M18	Brian May	Paul Stanley	63%	78%	95%	68%
M19	Tommy Lee	Joe Perry	52%	65%	90%	65%
M20	Jack Lemmon	Bing Crosby	67%	60%	88%	92%
M21	Iggy Pop	Anthony Kiedis	52%	62%	76%	60%
M22	Ted Nugent	Gregg Allman	64%	71%	95%	87%
M23	Richard Wright	John McIndoe	64%	51%	86%	70%
M24	Lou Reed	Bob Dylan	67%	64%	84%	73%

M25	Ray Davies	Steve Winwood	56%	47%	72%	80%
F01	Jane Wyman	Jane Wyatt	45%	60%	83%	75%
F02	Veronica Hamel	Tyne Daly	59%	56%	98%	92%
F03	Susan Sullivan	Susan Saint James	67%	76%	83%	85%
F04	Susan Hampsire	Patty Duke	69%	75%	90%	85%
F05	Susan Dey	Marlo Thomas	75%	69%	76%	95%
F06	Stephanie Powers	Anne Baxter	67%	67%	97%	80%
F07	Spring Byington	Cara Williams	45%	35%	81%	33%
F08	Shirley Booth	Irene Ryan	75%	73%	98%	80%
F09	Anna Magnani	Sada Thompson	81%	85%	100%	87%
F10	Imogen Coca	Carol Burnette	64%	65%	90%	72%
F11	Phyllis Kirk	Stockard Channing	63%	62%	90%	82%
F12	Angela Lansbury	Ann Francis	78%	67%	86%	95%
F13	Angie Dickinson	Peggy Wood	66%	45%	78%	88%
F14	Pat Benatar	Joan Jett	73%	64%	88%	78%
F15	Barbara Babcock	Diana Rigg	80%	55%	66%	90%
F16	Anne Meara	Mariette Hartley	56%	73%	91%	80%
F17	Barbara Bel Geddens	Joanne Woodward	58%	62%	76%	77%
F18	Barbara Parkins	Dianne Feinstein	55%	64%	83%	90%
F19	Jennifer Jones	Barbara Bain	61%	71%	81%	58%
F20	Brenda Vaccaro	Glenda Jackson	61%	69%	86%	77%
F21	Julie Harris	Melissa Anderson	59%	87%	95%	90%
F22	Lana Turner	Nancy Kelly	81%	73%	90%	83%
F23	Leslie Caron	Donna Reed	63%	69%	76%	78%
F24	Lynda George	Dinah Shore	61%	55%	74%	75%

F25	Michele Lee	Mary Tyler Moore	50%	62%	84%	67%
Testing						
M01	Alfred Lunt	Charles Boyer	77%	87%		
M02	Clark Gable	Cary Grant	59%	74%		
M03	James Cagney	Gene Hackman	77%	73%		
M04	Gary Cooper	Efram Zimbalast	75%	65%		
M05	Fredric March	James Garner	69%	75%		
M06	Fred Astaire	Alec Guinness	67%	83%		
M07	Gene Kelly	Marlon Brando	57%	50%		
M08	Basil Rathbone	John Howard	70%	77%		
M09	John Huston	Jackie Cooper	67%	80%		
M10	Buster Keaton	Burt Lancaster	78%	65%		
M11	Laurence Olivier	Lionel Barrymore	55%	63%		
M12	Roddy McDowell	Montgomery Clift	58%	65%		
M13	Paul Muni	Peter Falk	54%	77%		
M14	Kirk Douglas	Robert Donat	66%	82%		
M15	Frank Morgan	Leslie Nielson	58%	62%		
M16	Gregory Peck	Peter O'Toole	56%	61%		
M17	Richard Dix	Franchot Tone	61%	67%		
M18	Spencer Tracy	Jimmy Stewart	52%	68%		
M19	Charles Laughton	Orson Welles	84%	76%		
M20	William Holden	Tony Curtis	55%	54%		
F01	Audrey Hepburn	Elizabeth Taylor	71%	76%		
F02	Lauren Bacall	Ava Gardner	50%	67%		
F03	Fay Bainter	Barbara Stanwyck	49%	55%		
F04	Claudette Colbert	Betty White	60%	66%		

F05	Betty Davis	Joan Crawford	71%	40%
F06	Debbie Reynolds	Greer Garson	64%	75%
F07	Marlene Dietrich	Gretta Garbo	75%	58%
F08	Irene Dunne	Lynn Fontanne	75%	58%
F09	Mitzi Gaynor	Gladys George	66%	72%
F10	Ann Harding	Mary Pickford	39%	77%
F11	Helen Hayes	Carol Lombard	53%	54%
F12	Rita Hayworth	Judy Garland	52%	64%
F13	Ingrid Bergman	Luise Rainer	54%	52%
F14	Grace Kelly	Doris Day	60%	76%
F15	Piper Laurie	Lee Remick	66%	66%
F16	Norma Shearer	Shelly Winters	60%	68%
F17	Shirley Temple	Shirley MacLaine	57%	42%
F18	Deborah Kerr	Patricia Neal	52%	60%
F19	Natalie Wood	Anne Bancroft	63%	64%
F20	Rachel Roberts	Merlina Mercuri	84%	59%

Appendix C



UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

February 24, 2016

MEMORANDUM

TO:	James Lampinen
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT CONTINUATION
IRB Protocol #:	15-02-523
Protocol Title:	*Human Face Project: Perception and Memory of Faces*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Previous Approval Period:	Start Date: 03/02/2015 Expiration Date: 03/01/2016
New Expiration Date:	03/01/2017

Your request to extend the referenced protocol has been approved by the IRB. If at the end of this period you wish to continue the project, you must submit a request using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. Failure to obtain approval for a continuation on or prior to this new expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

This protocol has been approved for 5,800 total participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu