Science & Justice xxx (2016) xxx-xxx

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com

When age-progressed images are unreliable: The roles of external features and age range☆

William Blake Erickson[a, *, 1], James Michael Lampinen[a], Charlie D. Frowd[b, 2], Gregory Mahoney[c]

[a] *University of Arkansas, United States*
[b] *University of Winchester, United States*
[c] *Boston Police Department, United States*

ARTICLE INFO

Article history:
Received 1 May 2016
Received in revised form 20 November 2016
Accepted 29 November 2016
Available online xxx

Keywords:
Missing persons
Forensic imaging
Age progression
Reliability

ABSTRACT

When children go missing for many years, investigators commission age-progressed images from forensic artists to depict an updated appearance. These images have anecdotal success, and systematic research has found they lead to accurate recognition rates comparable to outdated photos. The present study examines the reliability of age progressions of the same individuals created by different artists. Eight artists first generated age progressions of eight targets across three age ranges. Eighty-five participants then evaluated the similarity of these images against other images depicting the same targets progressed at the same age ranges, viewing either whole faces or faces with external features concealed. Similarities were highest over shorter age ranges and when external features were concealed. Implications drawn from theory and application are discussed.

1. Introduction

Recovering missing persons is a major challenge for law enforcement organizations, especially when the missing person is a child. Although most missing children are recovered quickly (within 3 to 72 h), an important subset of cases see children missing for many years, and sometimes decades [10]. In these cases, helpful leads decrease exponentially as time progresses, substantially reducing the probability of the child's recovery [21]. One technique employed by members of law enforcement to generate new leads for these cases is to produce age-progressed images that approximate the appearance of the child at a later age. These images are then disseminated to the media so that the public can see them and potentially aid in the children's recovery.

Although techniques to create age progressions vary, the general method is for artists to incorporate knowledge of average craniofacial morphogenesis as well as images of the children's parents at various ages to narrow down the predicted appearance [32]. Any methods beyond this are idiosyncratic among artists [17]. Moreover, faces mature and age in generally predictable ways, so one might expect artists striving to create photorealistic images of a person's future appearance to create similar renditions. In the current paper we present a study aimed at determining the reliability of results produced by similar age-progression training on the age progressions themselves.

1.1. *Laboratory studies of age progressed images*

Rendering age-progressed images has some intuitive appeal and is widely used in missing persons' cases. However, systematic laboratory studies investigating how recognizable these images are have not produced promising results [4,16,19], finding that outdated images of the children lead to recognition of the older individual just as well as age progressions themselves.

In the first study investigating how well age-progressed images lead to correct identifications, Lampinen, Arnal, et al. [16] obtained childhood photographs of volunteers at ages 7 and 12, as well as biological relatives at these ages. Professional forensic artists then created age progressions of the 7-year old images up to age 12. Participants studied four outdated images, four current age 12 images, or age progressions and were then given a task to organize photos of 12-year old children into two teams. Within the set of images were the targets studied previously, and participants were told to indicate if they recognized any image from the study phase. Subsequently, participants viewed four lineups containing the study images as well as plausible decoys (foils). For both measures, current images led to correct identifications more often than age progressions and outdated images, which in turn did not reliably differ from one another. A second experiment modified this procedure by presenting outdated images alongside age progressions, which is more often the actual practice. Still, this condition did not differ from outdated images. It is perhaps worth mentioning that outdated images led in some cases to more recognitions than age progressions, although this benefit was not reliable.

? Funding: This work was supported by the National Science Foundation grant number 1155207.

* Corresponding author at: Department of Psychological Sciences, McAlester Hall 9A, Columbia, MO 65211-2500, United States.
Email address: ericksonw@missouri.edu (W.B. Erickson)
[1] Present address: University of Missouri, United States.
[2] Present address: University of Central Lancashire, United Kingdom.

http://dx.doi.org/10.1016/j.scijus.2016.11.006
1355-0306/© 2016 Published by Elsevier Ltd.

Lampinen et al. [19] extended this research using images from a real case where the child was recovered after many years, but still found the age progressions produced equivalent recognition to outdated images. Concerned that Lampinen and colleagues' paradigm might be too difficult for participants, Charman and Carol [4] conducted a series of more classical identification studies wherein participants merely viewed images and then attempted to identify the targets from lineups. Their results again showed that age-progressed images produce less correct identification than current and outdated images; in this case, however, age progressions actually produced increased false identifications. The authors interpreted this result as demonstrating that age-progressed images increase the number of plausible targets whose appearance they might match.

1.2. Challenges to predicting future appearances

In light of the research outlined in Section 1.1, it is clear that attempting to render an accurate prediction of someone's future appearance is a challenging task. Facial appearance can change in many different ways, increasing the amount of error in prediction. When discussing this sort of error, it is useful to consider validity and reliability. With age progressions, validity refers to whether a progression performs its intended purpose; namely, whether it is a plausible future likeness of an individual. The aforementioned experiments focused on the validity of age progression as a methodology—that is, do the resultant images match the actual appearance of the target individual? Generally, the research has found this to be the case, although matches are no more representative of current appearance than outdated images. Reliability refers to whether the process of age-progression produces the same results across different forensic artists. Of course, some variability is expected with any tool or form of measurement, and minimal error is acceptable. Extreme variability such that two age progressions of the same person are actually dissimilar from one another should be of concern to proponents of the method and might draw the reliability of the method into question. It is this latter matter of reliability between artists that is the focus of the current paper.

Because over 55% of missing persons cases go unresolved [23] and 85% of unidentified remains remain so [24], artists rarely receive feedback on how well their predictions matched the person's veridical appearance. Therefore, determining the causes of variability in a controlled setting could lead to more refined methods for the generation and implementation of age-progressed images. In Section 1.3.1 we outline some of what we believe are the major contributors of variance among age progressed images.

1.2.1. Natural and lifestyle factors over time

Growth patterns of human faces follow similar and predictable sequences of changes throughout life [1,9]. For this reason, forensic artists commonly take into account average developmental patterns. There is nonetheless individual difference in the exact timing and magnitude of the various changes that occur, which is why artists also typically use family members' images as a guide to how the child may appear at different age points. Beyond this, lifestyle and environmental factors such as diet, drug use, and sun exposure can incur unpredictable changes to a person's appearance [1]. These factors compound and interact with natural morphogenesis as the person remains missing for many years.

From a face space account [33,34], the myriad age-related facial changes would render faces further distinct from their younger iterations because they exaggerate shape and texture. Indeed, older faces do resemble caricatures of their younger selves [14]. So, although a forensic age progression is an informed prediction, it remains subject to error similar to a statistical point estimate within face space. Importantly, individual forensic artists may be biased to posit a similar or different prediction from other artists given the exact same starting materials.

Lampinen, Erickson, Frowd, and Mahoney [18] manipulated the age range artists were tasked with creating progressions, varying from ages 5 to 12 years, 5 to 20 years, and 12 to 20 years for eight individuals. Average age progressions were also generated by morphing the four progressions made for each individual at each age range. Participants then compared progressions and morphs in pairs alongside their intended targets' older ages and description-matched foils, rating their similarities on scales of 1 (extremely dissimilar) to 7 (extremely similar). Similarity ratings for images varied by artist, although shorter age ranges produced progressions more similar to targets than did longer age ranges. This is a sensible finding because age range would contribute to variability for the reasons stated earlier in this section – more time allows for more natural and lifestyle changes in facial appearance. Importantly, morphs produced more similar depictions than average ratings collapsed across all artists, likely because morphs minimized idiosyncratic features depicted by individual artists as in composite sketches [3]. This indirectly indicates that inter-artist reliability for images of the same targets negatively affects target similarity.

1.2.2. Hairstyle changes

A person's appearance can vary widely on a day-to-day basis as well. Of course, overt disguises by those not wishing to be recognized can hinder accuracy [29]. Researchers have found elsewhere in the forensic science and eyewitness memory literatures that changes in external features – especially hairstyle – can influence face perception [8] and even hinder it [2,22,28]. In the related literature examining forensic composites of wanted fugitives, reducing the likeness of a suspect's hairstyle has been found to reduce correct naming rates of the composites [12,11]. In this case, hairstyle is meant to match a witness's memory of a perpetrator. However, in the case of age-progressions, hairstyles are arbitrarily chosen by the forensic artist and different artists might choose very different styles for the same individual. So, the variability of age-progressions for the same target might originate in variability in hairstyle. For this reason, we examine the effect of external features such as hairstyle on perceptions of age-progressed images made by different artists in the current study.

1.2.3. Artist experience and training

A final contributor to age progressions' variability is that artists have different levels of training and experience. The effect of different levels of training on age progressions has never been examined before, but Davies and Little [6] explored the effect of expertise on forensic composites (i.e., images of perpetrators or abductees generated from an eyewitness's description). Police sketch artists and art students each sketched likenesses of six faces. Police sketch artists' images were rated as more faithful likenesses by independent judges who were blind to condition. So, specialized forensic training does produce more faithful likenesses in the composite domain. But does experience help predictions of future appearance?

Another factor related to experience is precise methodology: does the artist sketch by hand, use photographic manipulations via computer, or some combination of both? Frowd, Carson, et al. [11] compared naming rates of forensic composites made either via sketch or a number of computerized systems. Computerized systems, which produce more photorealistic images, were spontaneously correctly named by participants more often than sketches. However, sketches performed better than other methods when targets were average-looking. In the current study, we focused on human artists' performance creating age-progressed images because they are still almost exclusively conscripted by law enforcement and advocacy groups for these types of forensic images. Also, little is known about how their techniques (e.g., focusing on internal vs external features) affect image variability. We surveyed each artist participating for their style and methods. No two artists work identically or have identical training, so each artist in our study presents a unique intersection of experience and methodology for us to consider when examining variability among images.

1.3. The current study

In the current study, we recruited eight professional forensic artists with varying levels of experience, each making age progressions based on images provided by eight volunteers. Artists were assigned to create age progressions across three distinct age gaps ranging from seven and 15 years. In this way, we could assess variability of short- and long-term age progressions. Artists were free to generate the progressions however they typically do so in real cases, and therefore varied in terms of their precise methods, experience, and training in forensic imaging.

The key difference between this study and the studies previously described is that we did not ask independent judges to compare age-progressed images to their intended targets. Instead, we paired age progressions depicting the same targets at the same age ranges and asked participants to rate how similar these images appear to one another. Average participant ratings of similarity serve as a good estimate of proximity within face space [31], giving us an indication of how objectively distinctive the facial images are from one another. To assess the influence of external features on similarity judgments, some participants compared cropped images showing only internal facial features (e.g., eyes, nose, mouth).

Our main hypotheses are as follows: 1) Age progressions across the long range will be rated less similar than across the short ranges because there is greater idiosyncratic change due to maturation, 2) Age progressions across the 12 to 20 range will be more similar than the 5 to 12 range because 12 year olds normally already have an early adulthood appearance, and 3) face images with external features cropped will be rated as more similar to whole faces because artists' variable exaggerations of external shape and hairstyle affect similarity judgments.

2. Method

2.1. Participants

Eighty-five university undergraduates (females = 63) at the University of Arkansas participated in this study as an extra credit opportunity for a psychology class. Participants were 18 to 29 years of age ($M = 20.32$, $SD = 1.99$). The ethnic makeup was as follows: 85% were identified as Caucasian, 6% as Hispanic, 4% as African American, 1% as Asian, 1% as Pacific Islander, 3% as "Other" (1% identified as "Time Lord", the remaining declined to specify their ethnicity).

2.2. Materials

Eight Caucasian volunteers (females = 4) provided images of themselves at ages 5, 12, and 20 years and images of their parents at these ages in exchange for monetary compensation. We next recruited eight qualified professional forensic artists to create our age-progressed images. Training varied depending on program undergone, but always included a mixture of fine and forensic arts courses and courses instructing basic anatomy and physiology. Table 1 displays other relevant information collected when we surveyed the artists.

For each age progression, the artists were provided with a picture of the target individual at one of two start ages (5 or 12) and with parental images at the ages of 5, 12, and 20. Artists then age progressed the pictures from the start age (e.g., 5) to each of older target ages (e.g., 12 and 20) using whatever techniques he or she would use in an actual case (see Table 2). The only constraint on the artists was that they were asked to depict each individual from a full frontal view and with a neutral expression. This standardization allowed for subjective comparisons to be made without any influence of expression or pose. Artists were split into two groups (A and B) for counterbalancing purposes. Each artist was provided with pictures of four volunteers at age 5 and four of the volunteers at age 12, counterbalanced across artists. This resulted in a total of 12 images from each of eight artists for a total of 96 age-progressed images. Note that for any individual, at any starting age, age progressions were produced by four different artists. After producing the age progressions, artists were monetarily compensated and shown the targets' actual age 20 images upon request.

Table 1
Information about artists including survey responses regarding their training, experience, and recoveries attributed to their images.

Artist	Group	Organizational affiliation	Years of experience	Proportion of child progressions	Hours of training specifically in age progressions	Hours to create age progressions	# recoveries in last 5 years
1	B	Law enforcement	10 to 15	10%	5 to 10	4 to 5	3
2	A	Law enforcement	6 to 10	0%	40 +	3 to 4	2
3	B	College/university	15 to 20	50%	40 +	1 to 2	0
4	B	Law enforcement	30 +	10%	5 to 10	5 +	3
5	B	Other (non-profit)	1 to 3	90%	40 +	3 to 4	3
6	A	Law enforcement	1 to 3	90%	40 +	5 +	1
7	A	Law enforcement	3 to 6	0%	40 +	4 to 5	0
8	A	Law enforcement	10 to 15	0%	40 +	1 to 2	0

Note: Further information such as sex, race, age, and nationality of artists is not provided because the forensic art community is small and information that specific would effectively reveal their identities. They are also participants in this study, and we are obligated to maintain their confidentiality as we would any participants. Moreover, we are contractually obligated not to reveal this information.

Table 2
Artists' survey responses to specific techniques and information used to create age-progressed images.

Techniques and tools used	Artist							
	1	2	3	4	5	6	7	8
Growth norm database		x				x		
Personal growth norm knowledge	x		x	x	x	x	x	x
Biological relative photos at target age and last known photo	x	x	x	x	x	x	x	x
Lifestyle information	x	x	x	x	x		x	x
Medical information	x	x	x	x	x	x	x	x
Computer algorithms						x		
Hand sketches	x			x				x
Photoshop (or similar editing software)	x	x	x	x	x	x	x	x
Other (describe)			x[a]					

Note:

[a] "Geographical information".

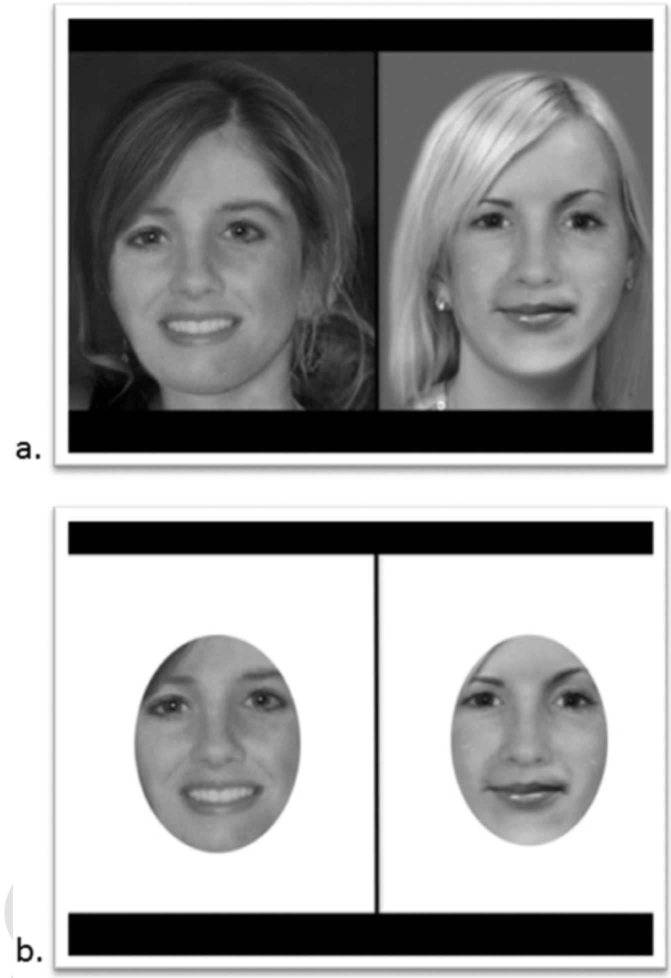

Fig. 1. Example comparisons of full (a) and cropped (b) age-progressions by two different artists in the 15-year age range.

2.3. Procedure

Participants took this study online using Qualtrics survey software loaded using an internet browser. They first answered demographic questions and then received instructions stating, "In this survey, you will see pairs of faces and rate how similar the face on the left is to the face on the right". They were shown an example of the Likert scale they were to use, where a rating of "1" is "extremely dissimilar", "4" is "neither similar nor dissimilar", and "7" is "extremely similar". They were also informed not to deliberate too long and to go with their first instinct when making decisions.

Participants then rated similarities of 144 image pairs (six possible pairwise comparisons multiplied by eight targets multiplied by three age ranges) presented in a random order. Each pair was displayed as a JPG image file sized 600 × 450 pixels where the two age-progressions were presented side-by-side (see Fig. 1). Every possible pairing of age-progressions of the same target and age range was created, appearing onscreen above the rating scale. Roughly half of the participants ($N = 43$ out of 85) were shown complete images side-by-side, and the remainder showed cropped age-progressions displaying only internal features. After rating similarities for all 144 images pairs, participants were debriefed, thanked, and provided with course credit.

3. Results

The purpose of this experiment was to measure the inter-artist reliability of professional forensic artists' age-progressed images of missing children. To do this, eight forensic artists created age-progressed images of eight targets across three age ranges. We then presented these images in same-target, same-range pairs of age progressions made by two different artists to participants who rated their similarity. In the following subsections we present the results of our analyses of these ratings.

3.1. Omnibus similarities and contrasts

We employed a mixed design for our main analysis, using a 2 (Face Type: Full or Cropped) × 3 (Age Range: 5 to 12, 5 to 20, and 12 to 20) design with Face Type as the between-subjects factor and Age Range as the within-subjects factor. There was no main effect of Face Type, $F(1, 83) = 2.36$, $p > 0.1$. However, a main effect of Age Range was found, $F(2, 166) = 535.72$, $p < 0.001$, $\eta^2_p = 0.87$,[3] $1 - \beta = 1.00$. Post-hoc pairwise comparisons revealed significant differences among all three age ranges, p's < 0.001, so that the 12 to 20 range ($M = 4.97$, $SD = 0.63$) yielded the most reliably similar images, followed by 5 to 12 ($M = 4.63$, $SD = 0.66$), which was more similar than 5 to 20 ($M = 3.71$, $SD = 0.76$). These means are displayed in Fig. 2.

Of particular interest was the significant Age Range x Face Type interaction, $F(2, 166) = 10.74$, $p < 0.001$, $\eta^2 p = 0.12$, $1 - \beta = 0.99$. Simple contrasts revealed that the interaction was driven by greater similarity of cropped faces than full faces in the 12 to 20 condition, $p = 0.02$, $\eta^2_p = 0.06$, and the 5 to 20 condition, $p = 0.048$, $\eta^2_p = 0.05$ (but not in the 5 to 12 condition, *ns*). So, external features have a great influence in older progressions, a finding which we consider in the Discussion.

3.2. Cross-race concerns

Comments on an early draft of this paper were that, given that our stimuli rendered only Caucasian faces, non-Caucasians may perceive

[3] Interpretation of effect sizes are taken from Cohen [5], such that small $\eta^2_p = 0.02$, medium $\eta_p{}^2 = 0.13$, and large $\eta^2_p = 0.26$.

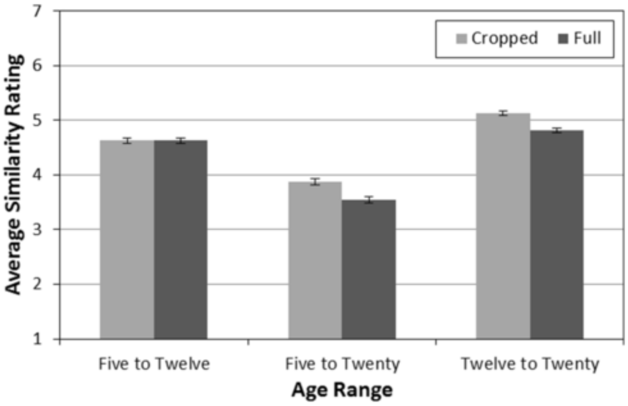

Fig. 2. Graph displaying the mean similarity ratings with standard error bars for each age range and face type.

them in a different way than Caucasians (e.g., featurally vs. holistically) and this could affect the interpretability of our results. The number of non-Caucasians ($N = 13$) was too small to directly compare responses by race, but we could estimate the effect by re-analyzing the data without these participants and observing whether the effects and patterns change.

We used the same mixed design for this analysis. This time, we found a main effect of Face Type, $F(1, 70) = 4.99$, $p = 0.029$, $\eta^2_p = 0.067$, $1 - \beta = 0.596$ such that cropped faces ($M = 4.58$, $SD = 0.66$) averaged higher similarity than full faces ($M = 4.24$, $SD = 0.70$). The main effect of Age Range was also found, $F(2, 140) = 480.85$, $p < 0.001$, $\eta^2_p = 0.87$, $1 - \beta = 1.00$ with 12 to 20 averaging highest similarity ($M = 4.96$, $SD = 0.65$), followed by 5 to 12 ($M = 4.60$, $SD = 0.67$) and 5 to 20 ($M = 3.69$, $SD = 0.77$). Post-hoc pairwise comparisons again revealed significant differences among all three of these age ranges, p's < 0.001. The Age Range x Face Type interaction was also obtained with this trimmed dataset, $F(2, 140) = 9.07$, $p < 0.001$, $\eta^2 p = 0.12$, $1 - \beta = 0.97$. Simple contrasts again found greater similarity of cropped faces than full faces in the 12 to 20 condition, $p < 0.01$, $\eta^2_p = 0.10$, and the 5 to 20 condition, $p < 0.01$, $\eta^2_p = 0.10$ (again not in the 5 to 12 condition, *ns*). That many of these effect sizes are larger than the previous analysis despite a smaller sample size and we obtained the Face Type main effect means that participant race may have affected similarity judgments.[4]

3.3. Difference from midpoint

Another way to examine the similarity ratings is to determine if they are significantly different from the midpoint (4) in our Likert Scale, which was labelled as "neither similar nor dissimilar". We analyzed age ranges as single-sample *t*-tests in separate analyses across Face Type and Age Range. For full faces, we found that all age ranges differed significantly from midpoint. The 5 to 12 range, $t(42) = 5.98$, $p < 0.001$, and 12 to 20 range, $t(42) = 8.75$, $p < 0.001$ were significantly above midpoint; the 5 to 20 range was significantly below, $t(42) = 3.93$, $p < 0.001$. This further indicates that the longer age range produces the greatest variability in progressions. For cropped faces, we found the 5 to 12 range, $t(41) = 6.42$, $p < 0.001$, and 12 to 20 range, $t(41) = 11.97$, $p < 0.001$, were significantly above midpoint but 5 to 20, $t(41) = 1.10$, *ns*, was not significantly below midpoint.

3.4. Relationship between validity and reliability

Our next analyses examine whether progressions' reliability relates to their validity. Put another way, we ask whether age progressions' similarities to their intended targets predicts their similarities to each other. To determine this, we created a target similarity index from the progression-to-target similarity ratings of these stimuli recorded previously [18] by adding each pair's member progressions' target similarity ratings together. We next calculated simple bivariate correlations between ratings and index scores for each image type (see Fig. 3). We found a strong significant positive correlation for full faces, $r = 0.48$, $p < 0.001$, and a moderate significant positive correlation for cropped faces, $r = 0.35$, $p < 0.05$, indicating that when age-progressed images are similar to each other, they are also similar to their intended targets.

3.5. Do differences between artists predict reliability?

An important set of questions arise relating to information found in Table 1. Namely, how do artists' years of experience, typical workload of child age progressions, and time taken to produce age progressions relate to reliability as we have defined it in this experiment? To answer this, we conducted a final set of correlations between image pair similarity ratings and the artist information described above.

The first correlation measured the relationship between image pair similarity ratings and the difference between each pairing's artists' respective years of experience. Because we asked artists to select their years of experience within 7 ranges of years, we first coded years of experience such that 1 was equal to 1–3 years of experience, 2 was equal to 3–6 years, 3 was equal to 6–10 years, 4 was equal to 10–15 years, 5 was equal to 15–20 years, 6 was equal to 20–30 years, and 7 was equal to > 30 years. Absolute differences in experience between artists were then obtained and correlated with average similarity ratings between images. We found a moderate significant negative correlation for full faces, $r = -0.41$, $p < 0.001$, and a moderate significant negative correlation for cropped faces, $r = -0.36$, $p < 0.001$, indicating that age-progressed images grow more disparate in appearance as gaps in experience between artists grow longer. Similar absolute differences between proportions of workloads devoted to age-progressions of children and hours spent generating age-progressed images were also calculated, but yielded correlations with image pairing similarity ratings near zero.

4. Discussion

We asked forensic artists to create age-progressed images of individuals across three age ranges and asked participants to rate how similar pairs of these images were to one another. This type of comparison differs from those found in previous investigations of age-progressed images in that it measures reliability, whereas previous investigations that compared similarity of age progressions to their intended targets measure validity. Taken together, our results demonstrate that variability was least pronounced at shorter age ranges and most pronounced at longer age ranges. Variability at two of the age ranges was due in part to differences in depicted external features. Additionally, images' similarity to each other is predicted by

[4] Another concern brought up by a reviewer was whether participant gender might affect similarity ratings. Although previous research has found no effect of participant gender in this type of paradigm, we investigated the possibility here. A series of exploratory analyses revealed no effects of participant gender on similarity ratings at any age range, t's < 0.5, p's > 0.05.

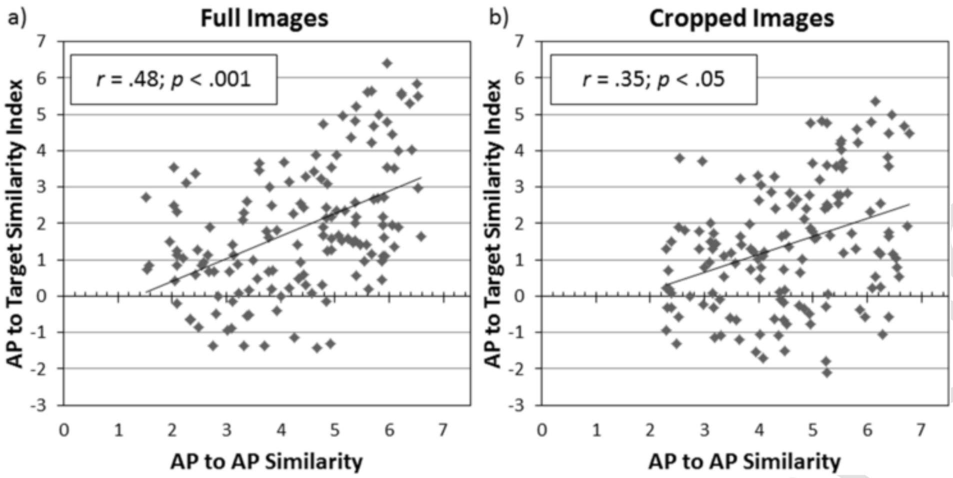

Fig. 3. Scatterplots with trend lines showing Target Similarity Index scores plotted against average AP to AP similarity ratings found in the current study for full (a) and cropped (b) age progressions.

how similar they are to their intended targets. In the following subsections we discuss these results in light of relevant theory and applications.

4.1. Why are predictions so disparate?

A major artist-level predictor for reliability of age-progressed images that we found was whether artists had comparable levels of experience. In particular, the farther the artists are apart in years of experience, the less likely their images are to be judged as similar to one another. However, this finding should be taken with caution because a previous study using these same stimuli found that artist experience had a weak relationship with similarity to targets [18]. Moreover, high similarity between images by artists at similar levels of experience could result from a cohort effect of a specific training experienced by generations of student-artists. Our other analyses revealed that remaining pertinent information gleaned from our surveys predicts no variability in images. Of interest is that amount of time spent generating age-progressed images and proportion of the person's workload consists of child of age-progressions did not predict reliability. Importantly, the variability among artists is adequately explained by specific techniques captured by our survey, which were very similar from artist to artist because they all used graphic design software and digital editing. These methodological similarities lead to the sensible prediction that age progressions would all be judged as fairly similar between artists, but this is not what we found.

Face perception theories predict that age progressions should be similar. Artists have superior visual expertise compared to laypeople, and artists' renderings are more accurate and detailed than those of non-artists [15,36]. Forensic artists in particular receive specialized training to render facial images and age them in photorealistic detail. Due to these similarities in techniques and expertise, one would think that age progressions made by different artists are all likely to fall in the same region of "face space" [34,35]. Face space is a multi-dimensional feature space that quantifies ways that face images vary. Such variability can be between people or between different images of the same person. Individual face images are located at specific coordinates within this space, and faces near one-another are rated as more similar than faces far apart in this space [31]. Given the same materials to work with, then, age-progressed images of the same person made by different artists should be near one-another in face space and therefore judged as similar-looking. However, we found that age-progressed images of the same individuals are not found to be particularly similar, suggesting that they are distant within face space.

Although no comparisons reached an average similarity rating of 6 (moderately similar), the only progressions that rated on the dissimilar side of the scale crossed the longest age range. Even these rated above a rating of 3 (somewhat dissimilar). So, images were never extremely dissimilar to one another, likely because images were all rendered partially or completely using digital editing of photographs portraying the actual targets' childhood appearances in combination of targets' parents and relatives. These faces would all reside near one another within face space. Variation then stemmed primarily from idiosyncratic choices made about targets' hairstyles and other external features, or because the predictions extended beyond ten years, both of which are what we found.

4.2. How to combat artist variability

How can age progressions be made more reliable? Forensic artists and scientists can draw practical solutions from the theoretical accounts laid out in Section 4.1. For example, software engineers have approached the problem of machine face recognition inspired in part by theories of human face recognition. Turk and Pentland [33] first developed a system called "eigenfaces", which runs principle-components analyses (PCA) on a set of standardized face images to capture a feature space very similar to Valentine's [35] face space model. An input face is then analyzed in comparison to these general feature dimensions and matched to a known exemplar. Automated systems using this approach have been successful and seen mainstream application [27], but facial aging can significantly reduce their reliability [28].

Recent work by Do, Homa, and Koehler [7] investigated how individuals' faces progress through face space when they age, because aging induces holistic changes (e.g., shape, texture, and complexion) within face space. They found that similar-looking (yet biologically unrelated) children undergo the same sorts of holistic transformations, resulting in adult faces that are also close together (i.e., similar-looking) within the space, and these similarities decline at older (> 60 years) ages. In other words, the trajectory a person's face takes through face space is predictable, so artists should therefore be able to generate images that are accurate likenesses if they use similar-looking individuals as guides. This in turn would minimize variability among artists.

An important discovery of the current paper is that progressions' validity (i.e., how similar they are to their intended targets, which has been the focus of previous investigations) predicts reliability. Images meant to resemble a common target should resemble one another, and this is the trend we found. However, the data could have revealed no sensible relationship or one in which age progressions resemble one another but not the targets they are intended to represent. Our results lead us to conclude that reliability improves as validity does. Therefore, the best way to increase reliability would be to develop techniques that create better likenesses.

Software engineering research has aimed to combat the aging problem specifically in the forensic context when a target is missing and has an outdated last known photograph [20,29]. Such a technique could potentially eliminate the need for an artist to be involved in the construction of age-progressions at all. Certainly, these algorithms produce believable likenesses of older countenances in a number of studies (e.g., [25,30]). Just as in the current study, investigators of computational approaches are limited by the number of high-resolution images depicting individuals over time, although several databases are freely available. The basic technique for generating age-progressed images via algorithm is to deform a younger face image, altering its texture and shape in such a way to create an older face, and then compare its new pixel values to those of a comparison image of an individual of the target age. Often such comparisons are also made by computers as well, as these techniques are most useful for automated security systems. Specific techniques vary and may be differentially optimal depending on circumstance (see [20] for an experimental comparison). Such approaches do produce believable likenesses over a lifespan of aging, but aesthetically pleasing likenesses are not necessarily accurate. Patterson and colleagues [26] address this limitation with an active appearance model (AAM) to create age progressions for a “ground truth” comparison to veridical images of the same targets at the intended ages. This is accomplished by actually comparing the pixel values of resultant images and standardized images of their intended targets. Using their newly assembled WARP image set, they made comparisons across many age gaps from 5 to 49 years. In addition, they examine gender and ethnicity non-specific models to models specified for these domains. As found with human artists, age progressions were better likenesses across shorter time spans. Gender specific models tended to be superior to gender non-specific models. Surprisingly, ethnicity nonspecific models performed better then ethnicity specific models, although the researchers concede this may have been due to inequities among ethnicity representations in the WARP image set. This research represents an important step for the development of algorithms, but the problem of capricious age-related changes remains, particularly for high frequency textural changes. Also unknown is whether these images are recognizable likenesses for human observers like those who might aid law enforcement in the field, as they have not been used in recognition memory paradigms or perceptual similarity rating studies.

We contend that the visual system of the human brain remains the best expert in perceiving, encoding, storing, and recognizing faces [37]. This means that forensic artists are able to visualize an older appearance in their minds well, but the appearance may get lost in translation during creation. To examine how idiosyncratic variations among artists affects accuracy, Lampinen et al. [18] compared the similarity of age-progressed images to their actual targets. They also created morphed average images of the progressions to minimize subtle images-level variation. These morphs were rated as similar to targets as the best age progressions. Because investigators in real cases cannot know which progression by which artist will be the best for a particular missing child, asking several artists to render images and morphing them in this way may be a promising alternative to the current single-artist procedure. Whatever the case, it is clear that the immediate future is one where artists and algorithms work together to create age progressions.

4.3. *Implications for other types of forensic imaging*

Some implications for the future of age progression as a technique extend from our findings. Clearly, it would be prudent for artists to maximize their attention on the missing person's internal facial feature configurations as their appearance and configuration change very little over time compared to external shape, which can conspicuously change during maturation and weight gain or loss. Likewise, hairstyle can affect identification accuracy while at the same time being something that can change capriciously. To combat this, artists could first generate the internal features and then produce multiple potential external feature variants. When law enforcement then disseminates these images, they could show several hairstyle or weight variations that could reasonably match the individual's veridical appearance.

Related to age-progression of potentially living missing persons is post-mortem depiction and facial reconstruction of victims whose remains have been recovered but are unidentifiable. Such instances are particularly common after natural disasters and terrorist attacks but also take place when remains are found in remote locations [38]. Because facial muscles are among the smallest in the body, they begin to deteriorate quickly after death compared to other muscles. The complexion takes on an ashen quality after blood flow stops, and various tissues swell and recede in turn, distorting the face of a newly deceased person in ways that can hinder identification based on facial appearance [32]. Over longer periods, faces exposed to the elements including insect or animal activity can be even further compromised. Like the anabolic and catabolic changes that take place in life, the general time course of post-mortem changes is well documented and post-mortem depiction is accomplished by turning the clock back after the interval since death is established. However, depending on environmental factors, the speed and extent to which changes manifest is variable. Like age-progression, then, the reconstruction is most accurate the sooner after post-mortem distortions begin. In most cases, hair actually tends to preserve well so there is less guesswork on the part of the artists in rendering these images. Also like age progressions, the area of post-mortem depiction suffers a dearth of systematic research and many cases involving recovered remains remain cold even when such images are created [23]. Moreover, missing children who are victims of kidnapping or human trafficking may age many years before death, which would further confound efforts to recreate their living visages. Clearly, both age-progression and post-mortem depictions suffer similar challenges that only further research manipulating timespan and specific techniques may help.

4.4. *Contributions and limitations*

Our major contributions – and the research questions we attempted to answer – are measuring how similar age progressions are when produced by different artists and determining the relationship between reliability and validity. This contrasts with previous research which has focused on validity as measured by recognition rates or similarity ratings to veridical images of respective targets. We also draw upon relevant theory, conceptualizing the images as they exist in a feature space to reveal that attempts to age faces based on the same starting materials can result in disparate areas of face space, and therefore regions less likely to increase recognition should the actual person be sighted. Finally, natural sources of variability such as the timespan over which the age progression predicts appearance and external feature variations such as hairstyle found that greater timespans produce more disparate appearance estimates and more variable external feature renderings.

A perceivable limitation to the current work is that some of our recruited artists do not regularly produce age progressions of children, and instead progress missing or fugitive adults. Artists who age-progress children fulltime may generate images more similar to one another because of their experience with that age group. However, we believe that our sample captures the natural variability found among forensic artists with varying levels of experience.

There a number of important variables related to the age progression problem that are not within the scope of the current paper but would be desirable to investigate. For example, a systematic investigation of how race of artists and targets interact to affect validity and reliability of age-progressed images is an area so far unexamined. This would be of particular relevance in the United States, where the majority of forensic artists are Caucasian whereas a disproportionate number of missing children are African American. People are generally better at perceiving and recognizing faces of their own race, so the question here is whether an artist might more accurately age-progress a child of his or her own race. The same question is valid for artist versus child biological sex. Future studies will address these concerns and explore further issues related to producing age-progressed images of missing children.

Uncited reference

[13]

References

[1] M. Albert, K. Ricanek, E. Patterson, A review of the literature on the aging adult skull and face: Implications for forensic science research and applications, Forensic Sci. Int. 172 (2007) 1–9, http://dx.doi.org/10.1016/j.forsciint.2007.03.015.

[2] V. Bruce, Z. Henderson, K. Greenwood, P.J.B. Hancock, A.M. Burton, P. Miller, Verification of face identities from images captured on video, J. Exp. Psychol. Appl. 5 (1999) 339–360, http://dx.doi.org/10.1037/1076-898X.5.4.339.

[3] V. Bruce, H. Ness, P.J.B. Hancock, C. Newman, J. Rarity, Four heads are better than one: combining face composites yields improvements in face likeness, J. Appl. Psychol. 87 (2002) 894–902, http://dx.doi.org/10.1037/0021-9010.87.5.894.

[4] S.D. Charman, R.N. Carol, Age-progressed images may harm recognition of missing children by increasing the number of plausible targets, J. Appl. Res. Mem. Cogn. 1 (2012) 171–178, http://dx.doi.org/10.1016/j.jarmac.2012.07.008.

[5] J. Cohen, Statistical power analysis for the behavioral sciences, Stat. Power Anal. Behav. Sci. 2nd (1988) 567, http://dx.doi.org/10.1234/12345678.

[6] G.M. Davies, M. Little, Drawing on memory: exploring the expertise of a police artist, Med. Sci. Law 30 (1990) 345–353.

[7] P. Do, D. Homa, K. Koehler, Identity categories and transformational paths for face changes across the age spectrum, Mem. Cogn. 42 (2014) 340–353, http://dx.doi.org/10.3758/s13421-013-0360-3.

[8] H.D. Ellis, J.W. Shepherd, G.M. Davies, Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition, Perception 8 (1979) 431–439, http://dx.doi.org/10.1068/p080431.

[9] D.H. Enlow, M.G. Hans, Essentials of Facial Growth, 1996http://dx.doi.org/10.1016/j.ajodo.2009.07.007.

[10] S. Finkelhor, H. Hammer, D. Schultz, National Estimates of Missing Children: An Overview, US Dept Justice, Off. Juv. Justice Delinq. Prev., 200212. https://www.ncjrs.gov/pdffiles1/ojjdp/196465.pdf.

[11] C.D. Frowd, D. Carson, H. Ness, J. Richardson, L. Morrison, S. McLanaghan, P. Hancock, A forensically valid comparison of facial composite systems, Psychol. Crime Law 11 (2003) 33–52, http://dx.doi.org/10.1080/10683160310001634313.

[12] C.D. Frowd, K. Herold, M. McDougall, L. Duckworth, A. Hassan, A. Riley, N. Butt, D. McCrae, C. Wilkinson, F. Skelton, A. McIntyre, P.J. Hancock, P. J. Hair today, gone tomorrow: holistic processing of facial-composite images, J. Exp. Psychol. Appl. (Manuscript submitted for publication).

[13] C.D. Frowd, F. Skelton, C. Atherton, M. Pitchford, G. Hepton, L. Holden, A.H. McIntyre, P.J.B. Hancock, Recovering faces from memory: the distracting influence of external facial features, J. Exp. Psychol. Appl. 18 (2012) 224–238, http://dx.doi.org/10.1037/a0027393.

[14] P.J.B. Hancock, A.C. Little, Adaptation may cause some of the face caricature effect, Perception 40 (2011) 317–322, http://dx.doi.org/10.1068/p6865.

[15] A. Kozbelt, Artists as experts in visual cognition, Vis. Cogn. 8 (2001) 705–723, http://dx.doi.org/10.1080/13506280042000090.

[16] J. Lampinen, J.D. Arnal, J. Adams, K. Courtney, J.L. Hicks, Forensic age progression and the search for missing children, Psychol. Crime Law 18 (2012) 405–415, http://dx.doi.org/10.1080/1068316X.2010.499873.

[17] J.M. Lampinen, W.B. Erickson, C.D. Frowd, C. D., G. Mahoney, Forensic Age Progression, in: S.J. Morewitz, C.S. Colls (Eds.), Handbook for Missing Persons, Springer, New York, (in press)

[18] J.M. Lampinen, W.B. Erickson, C.D. Frowd, G. Mahoney, Mighty Morphin' age progression: how artist, age range, and morphing influences the similarity of forensic age progressions to target individuals, Psychol. Crime Law 21 (2015) (Published Online).

[19] J.M. Lampinen, J.T. Miller, H. Dehon, Depicting the missing: prospective and retrospective person memory for age progressed images, Appl. Cogn. Psychol. 26 (2012) 167–173, http://dx.doi.org/10.1002/acp.1819.

[20] A. Lanitis, Comparative evaluation of automatic age-progression methodologies, EURASIP J. Adv. Signal Process. 2008 (2008) http://dx.doi.org/10.1155/2008/239480.

[21] I. McQueen, Computer age: computer enhanced aging, Police 33 (1989) 42–43.

[22] A. Memon, F. Gabbert, Unravelling the effects of sequential presentation in culprit-present lineups, Appl. Cogn. Psychol. 17 (2003) 703–714, http://dx.doi.org/10.1002/acp.909.

[23] National Missing and Unidentified Persons System, Missing Persons Database. https://www.findthemissing.org/en, 2015 (accessed 01.09.15)

[24] National Missing and Unidentified Persons System, Unidentified Persons Database. https://identifyus.org/en, 2015 (accessed 01.09.15)

[25] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, Comparison of synthetic face aging to age progression by forensic sketch artist, Proc. Seventh IASTED Int. Conf. Vis. Imaging, Image Process (2007) 247–252.

[26] E. Patterson, D. Simpson, A. Sethuram, Establishing a test set and initial comparisons for quantitatively evaluating synthetic age progression for adult aging. Biometrics (IJCB), 2014, IEEE International Joint Conf. on. IEEE (2014).

[27] P. Jonathon Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1090–1104, http://dx.doi.org/10.1109/34.879790.

[28] K. Ricanek, E. Boone, The effect of normal adult aging on standard PCA face recognition accuracy rates, In: Proc. Int. Jt. Conf. Neural Networks, 2005, pp. 2018–2023, http://dx.doi.org/10.1109/IJCNN.2005.1556210.

[29] G. Righi, J.J. Peissig, M.J. Tarr, Recognizing disguised faces, Vis. Cogn. 20 (2012) 143–169, http://dx.doi.org/10.1080/13506285.2012.654624.

[30] C.M. Scandrett, C.M. (nee Hill), C.J. Solomon, S.J. Gibson, A person-specific, rigorous aging model of the human face, Pattern Recogn. Lett. 27 (2006) 1776–1787, http://dx.doi.org/10.1016/j.patrec.2006.02.007.

[31] R. Scheuchenpflug, Predicting face similarity judgements with a computational model of face space, Acta Psychol. 100 (1999) 229–242, http://dx.doi.org/10.1016/s0001-6918(98)00019-5.

[32] K.T. Taylor, Forensic Art and Illustration, CRC Press, 2000.

[33] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991) 71–86, http://dx.doi.org/10.1162/jocn.1991.3.1.71.

[34] T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition, Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol. 43 (1991) 161–204, http://dx.doi.org/10.1080/14640749108400966.

[35] T. Valentine, Face-space models of face recognition, In: Comput. Geom. Process Perspect. Facial Cogn. Context. Challenges, 2001, pp. 83–113, http://dx.doi.org/10.1006/jmps.1997.1193.

[36] S. Vogt, S. Magnussen, Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen, Perception 36 (2007) 91–100, http://dx.doi.org/10.1068/p5262.

[37] G. Wallis, Toward a unified model of face and object recognition in the human visual system, Front. Psychol. 4 (2013) http://dx.doi.org/10.3389/fpsyg.2013.00497.

[38] C. Wilkinson, A. Tillotson, Post-mortem prediction of facial appearance, in: C. Wilkinson, C. Rynn (Eds.), Craniofacial Identification, Cambridge University Press, Cambridge, 2012.