University of Wollongong

Research Online

Faculty of Science, Medicine and Health -
Papers: Part B

Faculty of Science, Medicine and Health

1-1-2018

Variation and Likeness in Ambient Artistic Portraiture

Susan Hayes

University of Wollongong, suehayes@uow.edu.au

Nick Rheinberger

Australian Broadcasting Corporation

Meagan Powley

Red Point Artists Association

Tricia Rawnsley

Red Point Artists Association

Linda Brown

Red Point Artists Association

See next page for additional authors

Follow this and additional works at: <https://ro.uow.edu.au/smhpapers1>

Publication Details Citation

Hayes, S., Rheinberger, N., Powley, M., Rawnsley, T., Brown, L., Brown, M., Butler, K., Clarke, A., Crichton, S., Henderson, M., McCosker, H., Musgrave, A., Wilcock, J., Williams, D., Yeaman, K., Zaracostas, T., Taylor, A., & Wallace, G. G. (2018). Variation and Likeness in Ambient Artistic Portraiture. Faculty of Science, Medicine and Health - Papers: Part B. Retrieved from <https://ro.uow.edu.au/smhpapers1/28>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Variation and Likeness in Ambient Artistic Portraiture

Abstract

An artist-led exploration of portrait accuracy and likeness involved 12 Artists producing 12 portraits referencing a life-size 3D print of the same Sitter. The works were assessed during a public exhibition, and the resulting likeness assessments were compared to portrait accuracy as measured using geometric morphometrics (statistical shape analysis). Our results are that, independently of the assessors' prior familiarity with the Sitter's face, the likeness judgements tended to be higher for less morphologically accurate portraits. The two highest rated were the portrait that most exaggerated the Sitter's distinctive features, and a portrait that was a more accurate (but not the most accurate) depiction. In keeping with research showing photograph likeness assessments involve recognition, we found familiar assessors rated the two highest ranked portraits even higher than those with some or no familiarity. In contrast, those lacking prior familiarity with the Sitter's face showed greater favour for the portrait with the highest morphological accuracy, and therefore most likely engaged in face-matching with the exhibited 3D print. Furthermore, our research indicates that abstraction in portraiture may not enhance likeness, and we found that when our 12 highly diverse portraits were statistically averaged, this resulted in a portrait that is more morphologically accurate than any of the individual artworks comprising the average.

Publication Details

Hayes, S., Rheinberger, N., Powley, M., Rawnsley, T., Brown, L., Brown, M., Butler, K., Clarke, A., Crichton, S., Henderson, M., McCosker, H., Musgrave, A., Wilcock, J., Williams, D., Yeaman, K., Zaracostas, T. S., Taylor, A. C. & Wallace, G. (2018). Variation and Likeness in Ambient Artistic Portraiture. *Perception*, 47 (6), 585-607.

Authors

Susan Hayes, Nick Rheinberger, Meagan Powley, Tricia Rawnsley, Linda Brown, Malcolm Brown, Karen Butler, Ann Clarke, Stephen Crichton, Maggie Henderson, Helen McCosker, Ann Musgrave, Joyce Wilcock, Darren Williams, Karin Yeaman, T Zaracostas, Adam Taylor, and Gordon G. Wallace

Hayes, S., Rheinberger, N., Powley, M., Rawnsley, T., Brown, L., Brown, M., Butler, K., Clarke, A., Crichton, S., Henderson, M. and McCosker, H., 2018. Variation and Likeness in Ambient Artistic Portraiture. Perception, 47(6), pp.585-607. <https://doi.org/10.1177/0301006618770347>

PREPRINT

VARIATION AND LIKENESS IN AMBIENT ARTISTIC PORTRAITURE

Susan Hayes^{1,2*}, Nick Rheinberger³, Meagan Powley², Tricia Rawnsley², Linda Brown², Malcolm Brown², Karen Butler², Ann Clarke², Stephen Crichton², Maggie Henderson², Helen McCosker², Ann Musgrave², Joyce Wilcock², Darren Williams², Karin Yeaman², T.S. Zaracostas², Adam C. Taylor⁴, Gordon Wallace⁴

1. *Centre for Archaeological Science, University of Wollongong Australia*
2. *Red Point Artists' Association, Port Kembla, Australia*
3. *Australian Broadcasting Commission, ABC Radio Illawarra, Australia*
4. *ARC Centre of Excellence for Electromaterials Science, University of Wollongong Australia*

*corresponding author

ABSTRACT

An artist-led exploration of portrait accuracy and likeness involved 12 Artists producing 12 portraits referencing a 3D print of the same Sitter. The works were assessed during a public exhibition, and the resulting mean likeness assessments compared to portrait accuracy as indicated by geometric morphometrics (statistical shape analysis). Our results are that, independently of prior familiarity with the Sitter's face, the likeness judgements were higher for less morphologically accurate portraits. The two highest rated were the portrait that most exaggerated the Sitter's distinctive features, and a portrait that was a far more accurate (but not the most accurate) depiction. In keeping with research showing photograph likeness assessments involve recognition, we found familiar assessors rated the two highest ranked portraits even higher than those with some or no familiarity. Those lacking prior familiarity with the Sitter's face showed greater favour for the portrait with the highest morphological accuracy, and therefore most likely engaged in face-matching with the exhibited 3D print. Furthermore, our research indicates abstraction in portraiture does not enhance likeness, and in addition, we found that by averaging our 12 highly diverse artworks this results in a portrait that is more morphologically accurate than any of the 12 individual artworks.

INTRODUCTION

In August 2017, a public science project, *The Science of Portraiture*, was undertaken as part of Australia's National Science Week and involved the Red Point Artists Association (RPAA), the Australian Broadcasting Corporation's Radio Illawarra (ABC), and two of the University of Wollongong's research centres; the Centre for Archaeological Science (CAS) and the ARC Centre of Excellence for Electromaterials Science (ACES). The research question underpinning the project was the extent to which likeness judgements of artistic portraits undertaken and exhibited under relatively naturally occurring (ambient) conditions are related to their accuracy of depiction.

This project builds on previous research involving portrait drawings undertaken by one artist of 30 work colleagues, with each portrait being metrically and visually assessed for likeness (Hayes & Milne, 2011). This was the first study to see if geometric morphometrics (a landmark-based statistical shape analysis more typically applied to analyse variance in biological forms) could be usefully applied to artistic depictions. The results were that this approach can identify relative shape accuracy in artworks, but with regards to the influence of drawing accuracy on portrait likeness judgements the findings were generally inconclusive. Contributing factors were likely the relatively small number of assessors ($n=25$) with different levels of familiarity with the depicted, and that to control for this variation, the likeness assessments were laboratory-based, Power-Point displays of the portraits with the photographs from which they had been derived.

In contrast to the previous study, this current project is far more contextualised within actual art practices, and involved 12 Artists producing 12 unique portraits of the same Sitter, with each Artist achieving the parameters of the Sitter's face by referencing a 3D print. A public exhibition of the resulting artworks was attended by over 200 members of the general public, and 162 of these visitors chose to assess the portraits for likeness, with each recording, along with their likeness assessments, their level of prior familiarity with the Sitter's facial appearance.

Our concern with portrait accuracy is because, more than any other visual art genre, portraiture is understood to be mimetic, an artform that imitates life (von Alphen, 1997; West, 2004). Portraiture is also an unusual artform in that it is inescapably bound to the visual identity of a unique individual (Brilliant, 1991), to the extent that a portrait of low artistic merit will be judged as having greater authenticity if it is one of the few to have been undertaken from life (Barlow, 1997). As such, an exhibition of portraits is typically viewed "as a collection of people, rather than a display of art works" (West, 2004: 49). However, while art manuals strongly emphasise that accurately depicting a sitter's unique head pose,

feature shapes and their spatial configurations is essential within portraiture (e.g. Aristides, 2006; Edwards, 1999; Faigin, 1990; Maughan, 2004; Speed, 1917), the notion that portrait = person is not a literal equivalence. The use of the term 'likeness' carries with it an acknowledgement that portraits are similar – but not identical – to the person depicted (Brilliant, 1991). All portraits, whether they are sketched, painted, sculpted, photographed, scanned and/or 3D printed, are always imperfect copies. This is because, of necessity, the process of translating a person into a portrait involves partiality (not everything is included) and mediation (where what is included is transformed) (West, 2004). When this translation includes strategic absences that allow the viewer to actively see what isn't there, and transformations of distinctive features through exaggeration of these features, then a 'good' likeness is more likely to result (Arnheim, 1974; Gombrich, 1977, 1982; Ramachandran & Hirstein, 1999). And what constitutes a 'good' likeness in traditional portraiture is understood to be its capacity to convey immediacy, of eliciting a response in the viewer 'as if' the Sitter were present (Brilliant, 1991).

In addition to the active viewing that arises from filling in missing details and a 'peak shift' from the exaggeration of salient features (Ramachandran & Hirstein, 1999), it is likely that the 'immediacy' of a good portrait likeness includes its capacity to elicit recognition of the depicted. While the role of recognisability in portraiture is not specifically discussed in the literature, to some extent this can be inferred from the legal decision ~ 150 years ago that portrait likeness judgements can only be undertaken subjectively, and only by those intimately familiar with the appearance of the depicted (*Barnes v. Ingalls*, cited in Brilliant, 1991). Stronger support for the likeness judgements of portraits being related to their recognisability comes from face perception studies, where familiar face recognition is typically described as a rapidly processed perceptual Gestalt (i.e. immediacy), while unfamiliar viewers tend to engage in a piecemeal process of feature-by-feature comparison (for a recent review, see Young & Burton, 2017). Furthermore, and more centrally, a recent finding by Ritchie, Kramer and Burton (2018) is that likeness is, in fact, linked to recognition. This study involved ratings of "ambient" (naturally occurring and diverse) portrait photographs depicting the same person, and found that likeness judgements are significantly related to the viewer's prior experience with the facial appearance of the depicted, with the greater the level of prior familiarity, the higher the likeness ratings.

The use of "ambient images" in face perception studies is a relatively recent revision of experimental procedures that had tended to equate the perception of portrait photographs with actual faces, and an awareness that there is a high degree of variation across portrait photographs of the same person (Jenkins, White, Van Montfort, & Burton, 2011). This shift towards a greater use of naturally occurring (i.e. not purpose built), diverse images is also

linked to the notion that overly high levels of experimental control might lead to results that are more about the controls than the variable, and criticisms that researchers do not always differentiate between familiar face recognition and unfamiliar face matching, nor account for individual differences in the capacity to perform either task (Burton, 2013; Young & Burton, 2017). In effect, it is arguable that this revision has reframed photographs as what they are – artworks that are both partial and mediated translations of a person's face, and assessed for likeness according to how well they are recognised by familiars of the depicted.

Portrait photographs as artistic transformations is, of course, implicitly present in the face perception literature prior to the incorporation of ambient images, with transformative aspects being isolated and their effects studied extensively. This includes, but is not limited to, transformations arising from: head pose (Campbell, Benson, Wallace, Doesbergh, & Coleman, 1999), camera focal length (Třebický, Fialová, Kleisner, & Havlíček, 2016), image resolution (Bruce, Henderson, Newman, & Burton, 2001), texture and hue (Bruce & Young, 1998), lighting direction (Hill & Bruce, 1996; Johnston, Hill, & Carman, 1992), shearing (Hole, George, Eaves, & Razek, 2002), post-processing of feature size and position (Ge, Luo, Nishimura, & Lee, 2003; Hosie, Ellis, & Haig, 1988; Lewandowski & Pisula-Lewandowska, 2008), and automated exaggeration of distinctive features (caricaturing) (Benson & Perrett, 1991; Brennan, 1985; Rhodes, 1996). However, while these portrait photographs are artistic transformations, it has been argued that the use of predetermined variables results in alterations that are relatively simple and unilinear, and as such fall short of the actual diversity present in ambient portrait photographs (Ritchie & Burton, 2017). Nevertheless, compared to portrait drawings, even ambient photographs are relatively limited in what they transform.

Portrait drawings have also been studied, though one criticism is that they tend to be overly reliant on visual assessment (Ostrowsky, Cohen, & Kozbelt, 2014). Metric measures that have been applied to artworks include: geometry (Day & Davidenko, 2017; Perdreau & Cavanagh, 2014), spatial ratios/indices (Ostrowsky, 2015; Ostrowsky et al., 2014) and counts of stroke frequencies (Kozbelt, Seidel, ElBassiouny, Mark, & Owen, 2010). However, independently of whether the artworks are metrically or visually assessed, there is a marked tendency for the producers of the images to be either non-artists or a mix of non-artists and visual art students. Though there are a small number of studies that involve artists at work (e.g. Konecni, 1991; Miall & Tchalenko, 2001; Solso, 2001), this under-representation of artists in studies of artistic depiction is somewhat problematic, though no doubt due to the pragmatics of participant recruitment. A further potentially problematic area is that, whether or not artists are involved in the study, the images are both produced and assessed in ways that are

a-typical to art practices outside of the laboratory. In other words, as with most studies of portrait photographs, those involving portrait drawings tend to lack ambience.

To the best of our knowledge this is the first artist-led research project to explore artistic production, exhibition and likeness judgements under relatively ambient conditions, and one of the few to apply geometric morphometrics to analyse how, and to what extent, the shapes of the face are transformed when 12 Artists produce 12 portraits depicting the same individual (see Figure 1). Our area of particular interest is the extent to which likeness assessments are related to the relative accuracy of depiction and the assessor's prior familiarity with the Sitter's facial appearance.

Figure 1. The Portraits.

The alphanumeric code given to each portrait is indicated below

METHODS

Sitter and Artist Recruitment and Collaboration

Following the Sitter's agreement to participate, a call-out for RPAA members was undertaken. This resulted in 12 Artist collaborators, most of whom are highly experienced visual artists working across a range of media, though not all have extensive experience in portraiture. All were aware that participation in the project included the active and informed research collaboration of both Sitter and Artists regarding the study design, including identifying the research question, the conditions under which the portraits were produced and documented, the ways the resulting works were exhibited, how the works were assessed

and analysed, and the submission of the study as a co-authored research paper. Group input and agreement regarding these elements was achieved through a series of face to face meetings and email.

3D Print

To reduce variables regarding the Sitter's head-pose for the portrait, and to ensure the Artist-Sitter engagement was similar for all, each Artist set the parameters for their work in reference to a life-size 3D print of the Sitter's head and face. The Sitter was first digitally photographed at the ABC studios with his head in the standard anatomical position at a distance of 2.5m using a Canon EOS60D mounted on a tripod with the focal point of the Sitter's sellion (a point on the midline of the face at the deepest part of the nasal bridge below the brow). The Sitter's head and face was subsequently 3D scanned and printed at ACES who undertook this specifically for this project (see Supplementary Information).

Visual Support

None of the 12 Artists had prior familiarity with the Sitter's facial appearance, and therefore supplementary visual information was provided in preparation for the production of the portraits. This included links to footage of the ABC digital photography session (<https://www.facebook.com/abcillawarra/videos/1364291896949993>) and the ACES scanning session (<https://www.facebook.com/abcillawarra/videos/1402986829747166>), together with the digital photographs taken during both sessions (e.g. Figure 3). It was also agreed that each Artist could source additional images of the Sitter from the internet if they wished.

Figure 2. The ACES 3D print and the artist's easel orientation within the RPAA drawing studio.

(A) Vertical orientation of the 3D print and plinth, (B) Horizontal orientation of plinth and easel, (C) 3D print orientation (rotated 47° from facial midline).

Production of the Portraits

The 3D print produced by ACES was mounted and placed on an exhibition plinth (Figure 2a). The plinth was located in an RPAA drawing studio against a white wall, and a standing artists' easel was positioned 2.5m from the sellion of the unrotated 3D print (Figure 2b). The 3D print was then rotated 47° from the facial midline, resulting in the right cheek dominating (Figure 2c), and digitally photographed using a Nikon D3200 (f/5.6, exposure 1/125 sec., ISO-224, focal length 55 mm), again with the sellion as the focal point. Each Artist chose which leg of the standing easel was foremost when positioned at an oblique angle to the 3D print.

The Artists had 50 minutes to draft their portrait in reference to the 3D print, and each accomplished this using the artistic materials of their own choosing. On arrival at the RPAA drawing studio each Artist was provided with two sheets of 300gsm A2 mixed media paper on which to produce their portrait, which was to be unsigned. Prior to the commencement of their work referencing the 3D print, each Artist had their age, sex, height and handedness recorded (to identify if this had any impact on the resulting portrait) against a unique alphanumeric code, and once their time was complete each Artist photographed the 3D print using the same NikonD3200 camera and settings as detailed previously, in order to document the visual orientation they used to draw the portrait. The Artists, if they chose to do so, used their own cameras to take additional photographs of the 3D print to assist them in the completion of their portrait.

The Artists had two weeks to complete their portraits on one of the sheets of mixed media paper provided, and each chose the medium and scale of their work. The completed 12 portraits were delivered on a set day and time to the RPAA, and each portrait was given a different unique alphanumeric code by DA, a member of RPAA not directly involved in the research for this project. This was to facilitate the portraits remaining anonymous for all parties for the duration of the exhibition, the likeness judgements and all subsequent analyses. The 12 completed and anonymised portraits were then digitally photographed at a height of 1.5m using the Nikon D3200 camera (f/5.6, exposure 1/60 sec., ISO-110, focal length 32 mm) with the sellion as the focal point.

Geometric Morphometrics

A set of 19 homologous landmarks and 37 semi-landmarks (derived from curves) were identified for the photographs of the 3D print and the portraits (see Supplementary Information Figure 2 and Table 1). The 2D landmark coordinates (x,y) were digitized using the geometric morphometrics software *tpsDig232* (Rohlf, 2008, 2015). A different geometric morphometric program, *tpsSuper64* (Rohlf, 2015), was used to create two statistical

averages and the resulting x,y coordinates were exported. The first average involved the baseline photograph of the 3D print – see Figure 2c – and each of the 12 photographs of the 3D print taken by each of the 12 Artists. This average photograph of the 3D print, hereinafter referred to as the $\bar{X}$ 3D print, was used to assess the morphological accuracy of each of the portraits. This is because, by modifying the baseline photograph so as to be an average that includes the Artists' slightly different visual orientations, this reduced the potential of the artworks being assessed from a specific visual bias (the $\bar{X}$ 3D print is illustrated in SI Figure 2). The second statistical average was created from the 12 portraits, and is hereinafter referred to as the $\bar{X}$ Portrait.

The 12 individual Artist photographs of the 3D print, the 12 portraits, $\bar{X}$ 3D print, and the $\bar{X}$ Portrait, were analysed for morphological variance using the geometric morphometric software program, *morphologika2.5* (O'Higgins & Jones, 1998, 2006). Geometric morphometrics is a statistical shape analysis of landmark coordinates, and there is a statistical requirement that the number of specimens is greater than twice the number of landmarks if a multivariate regression (Goodall's F test) is undertaken (Cardini & Elton, 2007; Webster & Sheets, 2010). For this analysis the number of specimens (images) to landmarks (x,y coordinates) does not meet this statistical requirement and a multivariate regression is not feasible. Therefore, the focus was on identifying how the images cluster within the Principle Components (PC) capturing the greatest percentage of shape variance (e.g. PC1), and in identifying what are the major shape variances across the 12 portraits.

The Procrustes chord Distance arising from the Generalised Procrustes Analysis (GPA) performed by *morphologika2.5* was used to assess how morphologically accurate, overall, each of the portraits and the $\bar{X}$ Portrait are to the $\bar{X}$ 3D print. Undertaking a GPA to Procrustes register the landmarks is the first stage in a geometric morphometric analysis, and is where the set of landmark coordinates for each image are scaled, translated and rotated for comparative fit. Procrustes chord Distance is the overall distance of each image from each of the other images in shape space following Procrustes registration, and this distance (expressed as a numerical value) takes into account all shape variance, large and small, within the group. Therefore, plotting the Procrustes chord Distances (where the smallest distance is the closest morphological similarity) indicates how similar the 3D print photographs, $\bar{X}$ 3D print, the 12 portraits and $\bar{X}$ Portrait are to each other.

morphologika2.5 only reports statistical significance following a multivariate regression. Therefore, the Procrustes registered x,y coordinate landmarks were also entered into the statistics software *PAST3* (v.3.16) (Hammer, Harper, & Ryan, 2001) and a PCA performed (Bootstrap 1000; Scree plot and broken stick). The *PAST3* output includes which PCs, if any,

are statistically significant independently of a multivariate regression, and undertaking this supplementary analysis is enabled because the resulting eigenvalues and PC scores from the *PAST3* PCA agree, to 3 decimal places, with the *morphologika2.5* output.

More detailed explanations regarding how geometric morphometrics can be applied to analyse shape variance in artistic, forensic and archaeological images of the face can be found in the following: Hayes and Milne (2011) for frontal portraits, Hayes and Tullberg (2012) for witness descriptions of suspects, and Starbuck (2014) for morphological variance in archaeological figurines.

Anthropometrics and Facial Distinctiveness

Anthropometric measures and means of adult European male facial dimensions were sourced (George, 2007; Sforza et al., 2010; Sforza, Grandi, Binelli, et al., 2009; Sforza, Grandi, Catti, et al., 2009; Sforza, Grandi, De Menezes, Tartaglia, & Ferrario, 2011) and compared to the Sitter's facial dimensions. This was to identify the extent to which the Sitter's facial features are distinctive, and ascertain to what extent any distinctive features were retained and/or exaggerated in the portraits. The Sitter's facial data included measures taken during the photography session, which were also checked against the 3D print.

Exhibition

During a meeting to discuss the exhibition of the portraits, one of the Artists suggested the works be scaled and converted to greyscale in order to reduce these variables during the likeness assessments. Support for this suggestion was strong, but not unanimous. A compromise was agreed to where the works would be exhibited twice in two separate exhibition spaces – firstly as Greyscale prints and secondly as the Original artworks. To achieve the Greyscale prints the digital photographs of each portrait were loaded into the image manipulation software, *Adobe Photoshop CC 2017*, and each was scaled to the same facial height referencing the distance between the right exocanthion (outer eye corner) and stomion (midline point where the lips meet). Each portrait image layer had the levels adjusted using the automated levels function within the program, and all were cropped to the same size and converted to greyscale. A3 prints of the greyscale portrait photographs were produced by a local commercial printer, but printed in colour so as to retain the tonal variations of the artworks.

The Greyscale prints were mounted at a height of 1.5m in the RPAA Workshop Gallery. Each was clearly identified by the Artist's unique alphanumeric code, and exhibited in a randomised order. The 3D print was also displayed within the Workshop Gallery with the same height and orientation as was used for the initial production of the portraits (see Figure

2). The Original artworks were exhibited separately in the RPAA drawing studio, and while the Original artworks were also mounted on the studio walls at a height of 1.5m, they were exhibited in a different randomised order to that of the Greyscale prints.

Separate A4 Likeness Judgement recording sheets were produced for the Greyscale prints and Original artwork exhibitions. Judges were each allocated a unique numerical code, and asked to record their sex (male/female) and age group (<20, 20-35, 36-50, 51-65, >65). Following this were 5-point Likert scales with verbal low-high cues for familiarity with the Sitter's face (not at all - very well), the extent to which they would describe themselves as a visual artist (not at all - professional), and their assessment of the likeness of each portrait (very low - very high). The Greyscale and Original likeness judgements were presented in the same order as the Greyscale prints and Original artworks were respectively exhibited, and each likeness assessment had the alphanumeric code and a thumbnail of the corresponding portrait printed adjacent to the scale.

Figure 3. The Science of Portraiture exhibition promotional poster.

Likeness Judgements

The *Science of Portraiture* exhibition was promoted to the general public on the ABC radio, through the National Science Week, ABC and RPAA web pages, and by the distribution of posters and fliers to a range of local venues (Figure 3). The exhibition opened on the first day of National Science Week and ran for two weeks. Prior to the opening of the exhibition the Sitter attended a private viewing with SH (who was not one of the portrait Artists), assessed the Original artworks for likeness and selected the artwork he most preferred, independently of likeness (which was subsequently given to the Sitter as a gift).

Visitors to the exhibition who wished to participate in the likeness assessments were advised that what was meant by *likeness* was up to their interpretation (no further explanation was provided) and were encouraged to judge the Greyscale prints before judging the Original artworks (in which case the same unique identifier was used for each Judge). However, as a public science project, it was not compulsory for visitors to judge either or both versions of the 12 portraits. Out of over 200 visitors, 162 actively participated in the project by anonymously assessing the works (Judges): a small number ($n=7$) only assessed the Greyscale prints, 45 only assessed the Original artworks, and 110 assessed both the Greyscale prints and the Original artworks. Once completed, the Judges posted their assessment forms into a purpose-built, sealed carton, and collation of the assessment data was not undertaken until after the exhibition had closed.

Analyses of Judges and Likeness Judgements

In the instances where a Judge's assessment fell between the numerical boundaries of the Likert scale, these were recorded as half-way scores (e.g. 1.5, 2.5). The statistical picture of the cohort of Judges and the mean likeness judgements of the portraits were compiled using the following functions from the statistics software *PAST3*: descriptive statistics, bivariate correlation (Spearman's rho), cluster analysis (Ward's algorithm, Euclidian similarity index) and Mann-Whitney pairwise post-hoc tests with Bonferroni correction.

RESULTS

Geometric Morphometrics

Analysis of the photographs each Artist took of the 3D print and the artworks they produced indicates that the Artist's viewpoint, age, sex, and height did not significantly impact on the morphology of the 12 portraits. Although handedness did result in photographs of the 3D print that were more in profile for the left-handed Artists ($n=2$), this was not evident in the resulting portraits.

The geometric morphometric results of the PCA including the 12 portraits and the $\bar{X}$ 3D print is shown in Figure 4. In this analysis, 77% of the overall variance is captured by the first four Principle Components (PC1-4) with only PC1 attaining statistical significance (*PAST3*, Bootstrap 1000, Scree plot and broken stick). Figure 4 displays the plot of PC1 (x axis, 34% variance) and PC2 (y axis, 17% variance), and TPS (thin plate spline) wireframes extracted from the maximum specimen values (-0.07 and +0.07) of the PC1 axis. As can be seen in the shape variance displayed by the wireframes, the greatest variance is in the depiction of the Sitter's facial and nasal dimensions, which range across the portraits from long-thin to short-wide. The $\bar{X}$ 3D print sits within the positive values of PC1 and therefore tends towards short-wide in both facial and nasal shape. These characteristics are the most exaggerated in

Portrait AP117, which sits at the extreme positive values of PC1, and least present in Portrait AP111, which sits at the extreme negative values. Within the variance being captured by PC1, Portrait AP115 is most similar/closest in shape space to the $\bar{X}$ 3D print.

Figure 4. Geometric morphometric analysis of the 12 portraits and the $\bar{X}$ 3D print (open square).

Portrait codes are as indicated, and the portrait that has the lowest Procrustes Distance from the $\bar{X}$ 3D print is represented as a grey square. The TPS wireframes for PC2 are included for reference, but only the variance captured by PC1 attains statistical significance. The wireframes overlaid in the top right are from the maximum values on PC1, and clearly illustrates the morphological variance of the portraits in their depiction of the Sitter's facial and nasal dimensions.

While PC1 displays the statistically significant morphological variance, the Procrustes chord Distance (PD) scores account for all of the variance being captured by the coordinate landmark data entered into the geometric morphometric analysis. The PD scores attained by each of the portraits, from lowest (most similar) to highest (least similar), are as follows (and see Figure 5, x axis):

AP121 (0.049), AP123 (0.062), AP116 (0.065), AP118 (0.067)

AP115 (0.070), AP112 (0.071), AP122 (0.075), AP120 (0.081)

AP119 (0.082), AP117 (0.084), AP113 (0.103), AP111 (0.122)

As can be seen, Portrait AP121 (and not Portrait AP115) is closest to the overall morphological parameters of the $\bar{X}$ 3D print, and the portrait that PC1 indicates has most exaggerated the Sitter's nasal and facial dimensions (AP117) is amongst the portraits that are most dissimilar to the 3D Print and therefore least accurate.

When the 12 portraits are statistically averaged ($\bar{X}$ Portrait) and the resulting coordinate data entered into the PD analysis, the results are that the $\bar{X}$ Portrait supplants Portrait AP121 as being closest to the $\bar{X}$ 3D print (PD score 0.045). In other words, the statistical average of the 12 diverse portraits is the most accurate representation of the Sitter's facial morphoplogogy.

Figure 5. The $\bar{X}$ Portrait and Procrustes chord Distances

The $\bar{X}$ Portrait (left) is the average shape arising from the 12 portraits illustrated in Figure 1. The plot on the right is each of the 12 portraits (circles) ranked according to their Procrustes chord Distance (PD) from (i) the $\bar{X}$ Portrait (x axis), and (ii) from the $\bar{X}$ 3D Print (y axis). The portraits in bold/large circles are the portrait that is the most morphologically similar to the X3D print (AP121), and the portrait that is the most morphologically similar to the $\bar{X}$ Portrait (AP122).

Ranking the 12 portraits according to their Procrustes chord Distance from the $\bar{X}$ Portrait produces slightly different results. Although the $\bar{X}$ 3D print remains closest, morphologically, to $\bar{X}$ Portrait, it is Portrait AP122 (and not Portrait AP121) that is most similar to the statistical average of the portraits. The $\bar{X}$ Portrait is illustrated in Figure 5 (left), alongside a plot of each portrait according to the two sets of PD scores: (i) PD from the $\bar{X}$ Portrait (x axis) and PD from the $\bar{X}$ 3D print (y axis). PD is further discussed in the results reporting the overall relationship between morphological similarity, familiarity and likeness.

Sitter Facial Distinctiveness

Comparison of the sitter's facial feature dimensions with available published averages indicates the Sitter's facial features fall outside the normal range (are relatively distinctive) as follows:

- Wide nose relative to eye spacing and mouth width, but within the normal range in actual nose length and width (George, 2007); nasal width and length > 1 Standard Deviation (Sforza et al., 2011), though it should be noted that ‘normal’ for this anthropometric study is limited to 13 Italian males of similar age to the Sitter.
- Wide jaw relative to facial width (George, 2007).

Cohort of Judges

The 162 individuals (Judges) who assessed the portraits for likeness are skewed in sex, age and prior familiarity with the Sitter’s facial appearance: 68% are women ($n=110$), 62% are over 51 years in age ($n=100$) and 61% considered themselves unfamiliar with the Sitter’s face prior to attending the exhibition ($n=96$). Artistic experience is more evenly represented with half of the Judges being experienced artists (>3 on the Likert scale). Spearman’s rho indicates there is a slight tendency for the Judges with artistic experience to be female ($r = 0.18$, $p = 0.04$), and for the levels of artistic expertise to increase with age ($r = 0.22$, $p = 0.01$). There was no noticeably significant correlation (i.e. $p < 0.05$) between the Judges’ sex, age and artistic experience, and how they assessed the portrait likenesses.

Paired Portrait Judgements: Greyscale prints and Original artworks

Of the 162 Judges, 110 assessed both the Greyscale (cropped and scaled) prints and Original artworks. The profile of this sub-cohort of Judges is very similar to the total data set (65% female, 61% over 51 years, 61% unfamiliar with the Sitter’s face, 50% with artistic experience), and as with the larger cohort, there are no significant correlations (i.e. $p < 0.05$) between the sex, age and artistic experience of the Judges, nor, following Bonferroni adjustment, any significant impact of these individual differences on how the Judges assessed either the Greyscale prints or the Original artworks.

Across both exhibition conditions each of the portraits received at least one assessment indicating it was a poor likeness (≤ 1.5) and at least one assessment indicating a good likeness (≥ 3.5). However, a cluster analysis of the raw likeness data in *PAST3* (Ward’s algorithm, Euclidian similarity index), results in each of the portrait pairs forming a distinct cluster pair, and indicates a high degree of overall similarity between how the Greyscale prints and Original portraits were judged. There are differences in that 8 of the 12 portraits have a lower mean likeness as a Greyscale print, but this difference is only statistically significant for Portrait AP118 ($p = 0.03$, Mann-Whitney pairwise post-hoc test with Bonferroni correction). However, while significant, the impact is relatively negligible: Portrait AP118, which is already a monochromatic Original, is the second highest mean likeness in the Greyscale condition and highest mean likeness as an Original artwork. This movement does not change the order because Portrait AP117, which is a coloured artwork, has the highest mean likeness

rating in the Greyscale condition and is second highest within the Original artworks. A more noticeable result arising is one of similarity in the overall range of assessments – Portrait AP121 has the lowest level of agreement in the likeness judgements (highest standard error and highest variance) whether it is exhibited as a Greyscale print or in its Original monochromatic form.

Table 1. Mean, Variance and Standard Error (SE) of the Portrait Likeness by 110 Judges across both exhibition conditions (scaled Greyscale prints and Original)

Portrait	Greyscale prints			Original artworks		
	Mean	Variance	SE	Mean	Variance	SE
AP111	1.78	0.59	0.07	1.67	0.44	0.06
AP112	2.91	1.15	0.10	2.89	1.14	0.10
AP113	2.23	0.69	0.08	2.48	0.79	0.08
AP115	2.75	1.05	0.10	2.95	1.12	0.10
AP116	2.66	0.73	0.08	2.74	0.80	0.09
AP117	3.40	1.30	0.11	3.68	1.07	0.10
AP118	3.39	1.03	0.10	3.71	0.80	0.09
AP119	2.88	1.24	0.11	2.74	1.23	0.11
AP120	2.22	0.71	0.08	2.33	0.84	0.09
AP121	3.02	1.86	0.13	2.95	1.80	0.13
AP122	2.56	1.04	0.10	2.87	1.28	0.11
AP123	3.16	1.25	0.11	3.19	0.92	0.09

Judge Familiarity and Likeness Judgements

Given there is little real difference between how the Greyscale and Original artworks were assessed for likeness, and to avoid repetition yet maintain similar assessment conditions, the analyses of the likeness assessments involves only the likeness judgements of the Original artworks by those who reported their level of prior familiarity and assessed both the Greyscale and Original artworks (n=108).

Figure 6 shows the mean likeness assessments according to three groups based on the different levels of prior familiarity the Judges had with the Sitter's facial appearance (None = Likert scale 1-1.5; Some = Likert scale 2-2.5; Familiar = Likert scale 3-5). Although Spearman's rho (with Bonferroni correction) indicates there is general agreement between those with some or no familiarity ($r = 0.089$, $p = 0.001$) and those with some or higher levels of prior familiarity with the Sitter's face ($r = 0.77$, $p = 0.04$), there is little agreement between the extremes of familiarity (none and familiar, not significant). When the mean likeness

assessments according to familiarity level are compared to the individual portraits (Mann-Whitney pairwise post-hoc test with Bonferroni correction), for four of the artworks there is a significant difference in the assessments between those with no familiarity and those most familiar with the Sitter's facial appearance. The most familiar assessors tend to rate Portrait AP112 ($p = 0.04$), Portrait AP116 ($p = 0.03$), and Portrait AP121 ($p = 0.001$) a lower likeness. In contrast, while all of the Judges have tended to rate Portrait AP117 a good likeness, familiar assessors have, on average, given this artwork an even higher likeness rating ($p = 0.02$).

Figure 6. Comparison of Likeness Judgements by Level of Familiarity

The dashed line are the mean likeness judgements ($n=110$), while the bars show the level of prior familiarity the Judges had with the Sitter's facial appearance ($n=108$): no familiarity (white), some familiarity (mid-grey), moderate to high familiarity (dark grey).

An issue with these results is that due to the ambient composition of the cohort of Judges, twice as many assessments were undertaken by people with no prior familiarity with the Sitter's face (None: $n = 69$; Some: $n = 17$; Familiar: $n = 22$). It would be anticipated that lower levels of agreement would be evident in the 'Some' and 'Familiar' groups due to their relatively small numbers, but this does not hold true for most of the portraits. The descriptive statistics for the likeness judgements by level of familiarity are provided in Table 2, and while the Standard Error is predominantly higher for the smaller numbers of more familiar Judges, this is not the trend for the Variance or Standard Deviation. As can be seen, with the exception of Portrait AP112 and Portrait AP122, the Judges who had the greatest familiarity with the Sitter's facial appearance prior to assessing the artworks show a stronger level of within-group agreement than those lacking this foreknowledge. However, where all Judges appear to agree is in disagreement with the likeness displayed by Portrait AP121,

which has the highest Standard Deviation, Standard Error and Variance irrespective of the Judges' prior level of familiarity with the Sitter's facial appearance.

Sitter Likeness Assessments and Selection

The Sitter assessed the likeness of the Original portraits prior to the exhibition opening, and these results were not included in the calculations of the mean likeness assessments. The Sitter's assessments closely agree with the familiar assessor likeness means (Spearman's rho, $r = 0.83$, $p = 0.006$ with Bonferroni adjustment), but not the overall mean likeness or the assessments by those with some or no familiarity with his facial appearance. However, while the Sitter agreed with his familiars that Portrait AP117 is the best likeness, closely followed by Portrait AP118, and did not consider Portrait AP121 to be a good likeness, the Sitter also rated Portrait AP122 as a good likeness (equal rank to Portrait AP118), and it was Portrait AP122 that the Sitter selected, without hesitation, as the artwork he most preferred.

Table 2. Mean (X), Standard Error (SE), Variance (Var.) and Standard Deviation (SD) of the Portrait likeness assessments by level of prior familiarity (N=108)
Highest values for each statistic are shown in bold.

Portrait	No Prior Familiarity (n=67)				Some Familiarity (n=17)				Familiars (n=22)			
	X	Var.	SE	SD	X	Var.	SE	SD	X	Var.	SE	SD
AP111	1.8	0.44	0.08	0.67	1.6	0.65	0.20	0.81	1.5	0.27	0.11	0.52
AP112	3.1	1.00	0.12	1.00	2.6	1.14	0.26	1.07	2.4	1.36	0.25	1.17
AP113	2.6	0.78	0.11	0.88	2.4	0.82	0.22	0.91	2.1	0.69	0.18	0.83
AP115	3.1	1.17	0.13	1.08	2.7	0.85	0.22	0.92	2.5	0.93	0.21	0.96
AP116	2.9	0.80	0.11	0.90	2.6	0.79	0.22	0.89	2.3	0.56	0.16	0.75
AP117	3.5	1.07	0.12	1.03	4.1	0.98	0.24	0.99	4.1	0.72	0.18	0.85
AP118	3.6	0.87	0.11	0.93	3.9	0.59	0.19	0.77	3.9	0.81	0.19	0.90
AP119	2.8	1.40	0.14	1.18	2.7	0.66	0.20	0.81	2.3	1.04	0.22	1.02
AP120	2.5	0.79	0.11	0.89	2.1	0.84	0.22	0.92	2.0	0.70	0.18	0.84
AP121	3.2	1.68	0.16	1.30	3.2	1.57	0.30	1.25	2.0	1.43	0.25	1.20
AP122	2.8	1.19	0.13	1.09	3.1	1.52	0.30	1.23	2.8	1.42	0.25	1.19
AP123	3.1	0.78	0.11	0.88	3.7	1.28	0.27	1.13	2.9	0.75	0.19	0.87

Likeness, Familiarity and Morphological Accuracy

Comparison (Spearman's rho) of the portraits' Procrustes chord Distance (PD) scores from the $\bar{X}_3D$ print with the overall mean likeness judgements does not attain significance. When the PD scores are compared to the mean likeness assessments by their three levels of

familiarity (None, Some, Familiar), only those assessments that were undertaken by Judges with no prior familiarity with the Sitter's facial appearance attain significance ($r = -0.61$, $p = 0.03$). However, this likeness/accuracy relationship is due to the influence of an outlier in the PD scores, AP111, and when this portrait is removed from the analysis, so does all significance. When each of the portraits' PD from the $\bar{X}$ Portrait is analyzed against the overall and familiarity likeness means, AP111 is not an outlier and there is no significant relationship to any level of likeness assessment.

DISCUSSION

This research project involved the mean likeness judgements of 12 portraits depicting the same Sitter being compared to the extent to which each portrait agrees with what each of the 12 Artists used as the basis for their artwork – the morphology of a 3D print. Referencing a monochromatic 3D print (rather than the Sitter in life) was a new and challenging experience for all of the Artists, but it ensured the portrait production conditions were similar for all, and constrained the variance of each portrait regarding the depiction of the Sitter's head pose. Previous studies have found that even subtle variations in head pose are typically the main contributor to morphological variance when geometric morphometrics is applied (e.g. Hayes, 2016; Hayes & Milne, 2011; Hayes & Tullberg, 2012), is a confounding factor for automated facial identification (for a review, see Ding & Tao, 2015), and is similarly problematic for face perception studies involving metric assessments (Young & Burton, 2017).

As an ambient, public science project, some elements were unanticipated, but it is arguable that these have contributed to the relative richness of the results. Firstly, even though the research was undertaken with all participants contributing to, and being aware of, the research aims and methods (e.g. the Artists were not the objects of the study), the resulting portraits are highly diverse (Figure 1). A contributory factor could be that, as with most professional portrait commissions, none of the Artists had prior familiarity with the Sitter's facial appearance. However, what was probably more influential was that while each Artist had access to additional visual support materials, the extent to which this material was incorporated into the portrait varied. For example, the Artists who produced AP117 and AP122 accessed a wide range of images of the Sitter, while the Artist who produced AP121 completed their work solely in reference to the 3D print. Secondly, although we anticipated the exhibition visitors would have differing levels of familiarity with the Sitter's facial appearance, we did not anticipate quite so many attendees with no prior familiarity. As with the 12 Artists, nearly all who assessed the portraits knew of the Sitter, but for a large proportion of the visitors this was only in the context of being regular listeners to the Sitter's daily ABC Radio *Mornings* program. For these visitors there was a curiosity to see what the

Sitter looked like, and therefore their reference to the Sitter's actual facial appearance was limited to the information provided by the 3D print (Figure 2) and promotional poster (Figure 3), both of which were exhibited alongside the portrait Greyscale prints. However, despite the likeness judgements being dominated by unfamiliar assessors, age and sex were not found to be influential factors, and as with a previous study (Kozbelt et al., 2010), we found a high level of overall agreement across assessors with different levels of artistic expertise.

The Artists' decision to separately exhibit Greyscale versions of their portraits, and have these assessed prior to the Original artworks, was to reduce the influence of the portraits' variations in scale and hue and provide a more level playing field for the visitors' likeness assessments. The results arising from the 110 paired likeness assessments undertaken by visitors to the exhibition was that while the Greyscale portraits tended to be ranked lower in likeness, the only real difference was that the two highest ranked portraits for likeness (Portrait 117 and Portrait 118) exchanged first and second position when viewed as an Original artwork. Although the impact of hue and scale does not seem to have been previously tested with portraits, that the mean likeness assessment pairs are in close agreement is similar to the findings with portrait photographs – (i) that the impact of hue is negligible providing the tonal variation is maintained (Bruce & Young, 1998; Kramer, Manesi, Towler, Reynolds, & Burton, 2017), and (ii) that the impact of scale is likely negligible providing the horizontal or vertical spatial distribution patterns of the features are retained (Hole et al., 2002).

Because the likeness assessments were similar across the Greyscale and Original portrait conditions, only the Original assessments were used to compare familiarity and likeness with the morphological accuracy of the portraits. By restricting the assessments of the Original portraits to those visitors who had also assessed the Greyscale prints and included their level of familiarity with the Sitter's facial appearance ($n=108$), this meant that prior to undertaking their assessments all Judges had been exposed to the 3D print of the Sitter's face as well as the promotional material. This means, in effect, that all assessors had some exposure to the Sitter's facial appearance prior to assessing the Original artworks. However, it is not known the extent to which this exposure constituted a priming effect and provided the otherwise unfamiliar viewers with a mental representation of the Sitter's face. It has been found that exposure to a range of ambient photographs of the same person can enhance familiarity and therefore recognition (e.g. Andrews, Jenkins, Cursiter, & Burton, 2015; Ritchie & Burton, 2017), but for this study, while the images are highly diverse, their ambience is constrained in that all depict the same view of the Sitter's face.

Four portraits have resulted in being highly relevant to one or more of the core variables examined in this study: morphological accuracy as determined from geometric morphometrics (Principle Components analysis and Procrustes chord distances); the relationship of morphological accuracy to likeness assessments, and in particular the role of feature exaggeration; and, differences in likeness assessments between familiar and unfamiliar Judges. The relationship of each of these four portraits to these variables, and how these relate to the relevant literature is as follows:

AP121: morphological accuracy, unfamiliarity and face-matching

Portrait AP121 is the most morphologically accurate depiction of the 3D print, and the 3D print is a life size, accurate depiction of the Sitter's facial morphology. Furthermore, AP121 is a monochromatic artwork that only depicts the information provided by the 3D print of the Sitter's face (i.e. minimal hair and neck), and therefore is the portrait that most emulates the limitations of the 3D print, and is further limited in that the Sitter's eyes appear closed. Portrait AP121 is considered a better likeness by those with no prior familiarity with the Sitter's facial appearance, and this suggests that many of these unfamiliar Judges assessed the portraits through a process of face-matching – that is, by directly comparing each portrait element to the appearance of this element in the 3D print.

Although this has not been previously tested in portraiture, face-matching has been extensively studied with photographs (e.g. Bruce et al., 1999; Bruce et al., 2001; Burton, Wilson, Cowan, & Bruce, 1999; Clutterbuck & Johnston, 2004; Kemp, Caon, Howard, & Brooks, 2016; Megreya & Burton, 2006; Shepherd, Davies, & Ellis, 1981; Young, Hay, McWeeny, Flude, & Ellis, 1985). There is general agreement across these studies that, when faces are unfamiliar, correct identification predominantly involves a direct comparison of each of the facial features, and these comparisons are highly sensitive to, and disrupted by, even slight differences in the depicted individual's facial appearance (e.g. head pose, hairstyle, expression, lighting, focal length), though this disruption may be reduced when external features are masked (Kemp et al., 2016). Compared to the near perfect recognition rates of familiar faces depicted across highly diverse and highly degraded photographic images, unfamiliar face matching (also referred to as picture recognition (Young & Burton, 2017)) has a poor success/high error rate. However, as has been noted, people display considerable variance in their capacity to either face match or recognise faces from photographs (Burton, White, & McNeill, 2010; Kemp et al., 2016).

Individual variation in capacities for unfamiliar face matching and familiar face recognition may account for the high level of variation in the assessments of Portrait AP121 across all levels of familiarity. However, while varied in their responses, visitors who had a moderate to

high familiarity with the Sitter's face prior to attending the exhibition did not tend to judge portrait AP121 a good likeness, and the difference between familiar and unfamiliar assessors in the likeness ratings is significant. This suggests that had it also been assessed (and it was an unfortunate oversight for us to not include this assessment), familiar Judges would likely have considered the 3D print to be a relatively poor likeness, particularly given people have difficulty either recognising familiar faces, or determining the sex of an unfamiliar face, when viewing monochromatic surface scans (Bruce et al., 1993; Bruce et al., 1991).

AP117: enhanced likeness through caricature

Portrait AP117 is one of the least morphologically accurate portraits, with this inaccuracy including exaggeration of the Sitter's nasal and lower face widths (refer Figure 4). Portrait AP117 was also one of the two most highly rated portraits for likeness, with visitors having moderate to high prior familiarity with the Sitter's facial appearance tending to accord both of the most highly rated portraits even higher likeness scores. This finding is in agreement with a recent study of ambient portrait photographs (Ritchie et al., 2018), and for Portrait AP117 the increase was statistically significant.

Anthropometric studies of adult European males indicates the Sitter's nasal and jaw widths are relatively distinctive features (George, 2007; Sforza et al., 2011), though frontal widths do not necessarily correspond to facial depths. Exaggeration of salient features is generally known to be a key element of most figurative artworks (Gibson, 1971; Gombrich, 1977; Ramachandran & Hirstein, 1999), including cave-art (Cheyne, Meschino, & Smilek, 2009). Exaggeration of distinctive features is specifically understood to enhance likeness in drawings (Arnheim, 1974) and portraiture (Gombrich, 1982), and a computer derived 'caricature effect' has also been found to be a key element in enhancing recognition of photographs of familiar faces, and of line drawings derived from photographs (e.g. Benson & Perrett, 1991; Brennan, 1985; Rhodes, 1996). It is very likely, therefore, that Portrait AP117 achieved a significant increase in likeness ratings from those already familiar with the Sitter's facial appearance because it incorporated caricature of the Sitter's relatively distinctive frontal nasal and jaw widths into a $\frac{3}{4}$ view. Incorporation of different views of the facial features into one view (e.g. increasing the profile of the nasal bridge) is one of the known transformations of portraiture (Sturgis, 1998), and its application is because it enhances the recognisability of the resulting face.

AP118: morphological accuracy and likeness

Portrait AP118 is the fourth most accurate portrait, attained the highest overall mean likeness score, is the highest mean ranked portrait for those with no prior familiarity with the Sitter's face, and second highest for those with some to a high level of familiarity,

including the Sitter himself. In many ways, the high ranking of Portrait AP118 contradicts the conclusions drawn from the two preceding artworks. Portrait AP118 has high morphological accuracy, with the geometric morphometric PC1 result indicating it has not exaggerated the Sitter's distinctive jaw and nasal widths (or at least not exaggerated those characteristics that were measured). Portrait AP118 is also a monochromatic artwork, and while there was no significant difference in the likeness assessments of this work as either a Greyscale or Original portrait, Portrait AP118 is somewhat visually similar to the most morphologically accurate portrait, Portrait AP121. However, unlike Portrait AP121 (and by implication the 3D print), Portrait AP118 was rated a good likeness by familiar 'recognisers' as well as unfamiliar 'face-matchers', and, as with the less accurate/caricatured Portrait AP117, experienced a 'peak' in likeness ratings by those most familiar with the Sitter's face prior to attending the exhibition.

What this outcome indicates is that accuracy of depiction does have a role in familiar as well as unfamiliar likeness judgements of portraits, and that it is possible that caricature does not have to be strongly present to elicit higher likeness ratings by familiar assessors. This finding is supported to a limited extent by a study of accuracy in face drawings by non-artists (Ostrowsky et al., 2014), which found a weak-moderate relationship between metric accuracy and visual assessments. But what is most missing from our study of metric accuracy for all of the portraits, not just Portrait AP118, is the depiction of facial texture.

Facial texture has been found to play a dominant role in the recognisability of portrait photographs (e.g. Hancock, Burton, & Bruce, 1996), and it is highly probable that a relatively accurate and/or caricatured depiction of facial texture has enhanced the likeness assessments of Portrait AP118 and/or Portrait AP117. Software has been developed to statistically analyse facial texture in portrait photographs (Kramer, Jenkins, & Burton, 2017), and this software would no doubt also work with other portrait image types. Unfortunately, the application (*Interface*) requires fairly frontally orientated representations of the depicted face (e.g. both pupils visible), and as relatively strong $\frac{3}{4}$ views that are closer to being profiles, our portraits do not meet this precondition.

AP122: average diversity

Portrait AP122 attained only moderate mean likeness ratings across all levels of visitor prior familiarity. Portrait AP122 was, however, assessed as equally second best for likeness by the Sitter (alongside Portrait AP118). This portrait is relatively abstract (in that essential information is depicted with minimal detail), and was produced, as with Portrait AP117, in reference to a large number of additional images of the Sitter. Portrait AP122 is also the portrait the Sitter most preferred.

A study involving professional and amateur portrait artists has found that artists are more capable of abstraction when drawing from life – as opposed to drawing from memory – and that abstraction correlates positively with visual assessments of how interesting and pleasing a portrait drawing, with pleasing and interesting also being positively correlated (Konecni, 1991). This study, however, was assessed for these characteristics by those unfamiliar with the depicted, and did not, therefore, include likeness or recognition. If, however, abstraction (interesting/pleasing) is positively correlated with recognisability, it would be anticipated that Portrait AP122 would attract higher likeness ratings by familiars, but this did not occur.

As well as being stylistically interesting, Portrait AP122 is also interesting because it is the closest, morphologically, to the statistical average of all of the 12 artworks. Other than a similarity of head-pose and a commonality of art paper used, the 12 portraits the Artists produced are highly diverse. It was an unexpected outcome, therefore, to discover that our 12 divergent depictions of the Sitter's face and features result in a statistically average portrait that is the most morphologically accurate depiction of the Sitter's features – i.e. more morphologically accurate than AP121 (see Figure 8). Face perception studies have found that the average of divergent photographs of the same face, through their removal of aspects that are irrelevant to facial identity, are highly recognisable for familiar viewers (Young & Burton, 2017). However, while recognisable, compared to individual instances of ambient portrait photographs, facial averages have also been found to attract comparatively low likeness ratings by familiars (Ritchie et al., 2018).

This description of portrait photograph averages being recognisable because of their removal of irrelevant detail is similar to that of abstraction, where an artist selects and depicts essential information with minimal detail (Konecni, 1991). Abstraction is, arguably, also an extension of the inherent 'partiality' (Gombrich, 1977; West, 2004) of all artistic depiction – it is not possible for any artwork to replicate life, and portraits are not people. Portrait AP122 is a relatively abstract depiction of the Sitter's face that is similar, morphologically, to the average of the portraits, and not the average of portrait photographs of the Sitter's face. So, while the Sitter's selection of Portrait AP122 as the one he most preferred is likely related to this portrait being the most pleasing and most interesting, it is an indirect link to further suggest that Portrait AP122 was not rated highly for likeness because it shares characteristics of average photographs. Were this a more direct link, it would be possible to conclude that abstraction does not enhance likeness and recognisability in portraiture.

With respect to how the shapes of a Sitter's face are depicted in portraiture, and how this relates to likeness assessments, our research findings agree with previous research as follows:

- likeness assessments are relatively independent of artistic experience, which agrees with a previous study involving visual assessments of accuracy of referencing portraits traced from photographs (Kozbelt et al., 2010);
- colour and scale do not significantly influence likeness assessments of portraits, a finding that is similar to previous research with portrait photograph recognition (Bruce & Young, 1998; Hole et al., 2002);
- familiar viewers rate good likenesses higher than unfamiliar viewers, indicating likeness assessments in ambient portraiture are linked to recognisability, a finding that agrees with recent research involving ambient portrait photographs (Ritchie et al., 2018);
- unfamiliar viewers show a stronger preference than familiar viewers for more accurate portrait depictions, and therefore appear to engage in face-matching to ascertain likeness, which has been found to occur with unfamiliar viewers of portrait photographs (Young & Burton, 2017);
- exaggeration of distinctive facial features enhances a portrait's likeness for all viewers, and in particular familiar viewers, which is in agreement with aesthetic, psychological and neurological theories of art (Arnheim, 1974; Brennan, 1985; Gombrich, 1977, 1982; Ramachandran & Hirstein, 1999), as well as face perception studies referencing automated caricatures derived from photographs (e.g. Benson & Perrett, 1991; Rhodes, 1996);
- exaggeration of distinctive features does not need to be present for a good likeness to result, which agrees with previous findings of a weak-moderate positive correlation between metric accuracy and subjective likeness assessments of portraits produced by non-artists (Ostrowsky et al., 2014).

Furthermore, our research indicates:

- abstraction does not appear to enhance likeness/recognition in portraiture;
- highly diverse portraits of a Sitter's face, when statistically averaged, produce a highly accurate depiction;
- meaningful research can be undertaken under ambient conditions that are closer to the everyday of visual art practices (image production, exhibition and assessor participation).

ACKNOWLEDGEMENTS

The authors would like to thank the following people for their support and assistance with this project: Dulcie Dal Molin, Naomi Arrowsmith and Donna Abbati; Gert van den Bergh, Treva Taylor and Brett Powley; Steve White and Justin Huntsdale; Stephen Beirne and Chris Richards. The authors would also like to acknowledge the enthusiastic support of the visitors to the *Science of Portraiture* exhibition during National Science Week 2017, and for the encouragement and insightful recommendations provided by our two reviewers.

DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The portrait production, exhibition and analysis were self-funded by the RPAA membership; the scanning and 3D print was undertaken by ACES and the Australian National Fabrication Facility Materials Node, and donated to the project as part of National Science Week 2017.

REFERENCES

- Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology*, 68(10), 2041-2050.
- Aristides, J. (2006). *Classical Drawing Atelier: A Contemporary Guide to Traditional Studio Practice*. New York: Watson-Guption.
- Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley and Los Angeles: University of California Press.
- Barlow, P. (1997). Facing the past and present: the National Portrait Gallery and the search for 'authentic' portraiture. In J. Woodall (Ed.), *Portraiture: Facing the subject*. Manchester: Manchester University Press.
- Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1), 105-135.
- Brennan, S. E. (1985). Caricature generator: the dynamic exaggeration of faces by computer. *Leonardo*, 18(3), 170-178.
- Brilliant, R. (1991). *Portraiture*. London: Reaktion Books.
- Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., . . . Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, 22, 131-152.
- Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., & Linney, A. (1991). Recognising facial surfaces. *Perception*, 20(6), 755-769.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339.
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7(3), 207-218.
- Bruce, V., & Young, A. (1998). *In the Eye of the Beholder: the Science of Face Perception*. Oxford: Oxford University Press.
- Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, 66(8), 1467-1485.
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42(1), 286-291.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition from poor quality video: evidence from security surveillance. *Psychological Science*, 10(3), 243-248.
- Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: how poses that change brow position affect perceptions of gender. *Perception*, 28(4), 489-504.
- Cardini, A., & Elton, S. (2007). Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology*, 126(2), 121-134.
- Cheyne, J. A., Meschino, L., & Smilek, D. (2009). Caricature and contrast in the Upper Palaeolithic: Morphometric evidence from cave art. *Perception*, 38(1), 100-108.
- Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, 11(7), 857-869.
- Day, J. A., & Davidenko, N. (2017). Physical and perceptual accuracy of upright and inverted face drawings. *Visual Cognition*, 1-11.
- Ding, C., & Tao, D. (2015). A Comprehensive Survey on Pose-Invariant Face Recognition. *arXiv preprint arXiv:1502.04383*.
- Edwards, B. (1999). *The New Drawing on the Right Side of the Brain*. New York: Penguin Putnam.
- Faigin, G. (1990). *The Artist's Complete Guide to Facial Expression*. New York: Watson-Guption Publications.
- Ge, L., Luo, J., Nishimura, M., & Lee, K. (2003). The lasting impression of Chairman Mao: hyperfidelity of familiar-face memory. *Perception*, 32, 601-614.

- George, R. M. (2007). *Facial Geometry: Graphic Facial Analysis for Forensic Artists*. Springfield Illinois: Charles C Thomas.
- Gibson, J. J. (1971). The information available in pictures. *Leonardo*, 4, 27-35.
- Gombrich, E. H. (1977). *Art and Illusion: A Study in the Psychology of Pictorial Representation* (5th ed.). Oxford: Phaidon Press.
- Gombrich, E. H. (1982). The mask and the face: the perception of physiognomic likeness in life and art *The Image and the Eye: Further studies in the psychology of pictorial representation* (pp. 105-136). Oxford: Phaidon.
- Hammer, Ø., Harper, D., & Ryan, P. (2001). PAST-Palaeontological statistics. *www.uv.es/~pardomv/pe/2001_1/past/pastprog/past.pdf*, acessado em, 25(07), 2009.
- Hancock, P. J., Burton, A. M., & Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory and Cognition*, 24(1), 21-40.
- Hayes, S. (2016). A geometric morphometric evaluation of the Belanglo 'Angel' facial approximation. *Forensic Science International*, 268(Supplement C), e1-e12. doi:<https://doi.org/10.1016/j.forsciint.2016.09.009>
- Hayes, S., & Milne, N. (2011). What's wrong with this picture? An experiment in quantifying accuracy in 2D portrait drawing. *Visual Communication*, 10(2), 149-174.
- Hayes, S., & Tullberg, C. (2012). Police witness identification images: a geometric morphometric analysis. *Journal of Forensic Sciences*, 57(6), 1487-1494. doi:10.1111/j.1556-4029.2012.02168.x
- Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 986-1004.
- Hole, G. J., George, P., Eaves, K., & Razek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31(10), 1221.
- Hosie, J. A., Ellis, H. D., & Haig, N. D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception*, 17(4), 461.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313-323.
- Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21(11), 365-375.
- Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving Unfamiliar Face Matching by Masking the External Facial Features. *Applied Cognitive Psychology*, 30(4), 622-627.
- Konecni, V. J. (1991). Portraiture: an experimental study of the creative process. *Leonardo*, 24(3), 325-328.
- Kozbelt, A., Seidel, A., ElBassiouny, A., Mark, Y., & Owen, D. R. (2010). Visual selection contributes to artists' advantages in realistic drawing. *Psychology of Aesthetics, Creativity, and the Arts*, 4(2), 93.
- Kramer, R. S., Jenkins, R., & Burton, A. M. (2017). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior Research Methods*, 49(6), 2002-2011.
- Kramer, R. S., Manesi, Z., Towler, A., Reynolds, M. G., & Burton, A. M. (2017). Familiarity and within-person facial variability: The importance of the internal and external features. *Perception*, 0301006617725242.
- Lewandowski, Z., & Pisula-Lewandowska, A. (2008). The influence of change in the size of face elements on the perception of a woman's portrait. *HOMO-Journal of Comparative Human Biology*, 59(3), 253-260.
- Maughan, W. L. (2004). *The Artist's Complete Guide to Drawing the Head*. Watson-Guptill: New York.
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory and Cognition*, 34(4), 865.
- Miall, R. C., & Tchalenko, J. (2001). A painter's eye movements: a study of eye and hand movement during portrait drawing. *Leonardo*, 34(1), 35-40.
- O'Higgins, P., & Jones, N. (1998). Facial growth in *Cercopithecus torquatus*: An application of three dimensional geometric morphometric techniques to the study of morphological variation. *Journal of Anatomy*, 193, 251-272.

- O'Higgins, P., & Jones, N. (2006). Tools for statistical shape analysis. Retrieved from <http://hymns.fme.googlepages.com/resources>
- Ostrofsky, J. (2015). Developmental and geographic analyses of spatial biases in face drawings produced by children. *Empirical Studies of the Arts*, 33(1), 3-17.
- Ostrofsky, J., Cohen, D. J., & Kozbelt, A. (2014). Objective versus subjective measures of face-drawing accuracy and their relations with perceptual constancies. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 486.
- Perdreau, F., & Cavanagh, P. (2014). Drawing skill is related to the efficiency of encoding object structure. *i-Perception*, 5(2), 101-119.
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7), 15-51.
- Rhodes, G. (1996). *Superportraits: caricatures and recognition*. Hove: Psychology Press.
- Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *The Quarterly Journal of Experimental Psychology*, 70(5), 897-905.
- Ritchie, K. L., Kramer, R. S., & Burton, A. M. (2018). What makes a face photo a 'good likeness'? *Cognition*, 170, 1-8.
- Rohlf, F. J. (2008). tpsDig (Version 2.12): Department of Ecology and Evolution, State University of New York at Stony Brook.
- Rohlf, F. J. (2015). The tps series of software. *Hystrix, the Italian Journal of Mammalogy*, 26(1), 9-12.
- Sforza, C., Grandi, G., Binelli, M., Dolci, C., De Menezes, M., & Ferrario, V. F. (2010). Age- and sex-related changes in three-dimensional lip morphology. *Forensic Science International*, 200(1), 182. e181-182. e187.
- Sforza, C., Grandi, G., Binelli, M., Tommasi, D. G., Rosati, R., & Ferrario, V. F. (2009). Age- and sex-related changes in the normal human ear. *Forensic Science International*, 187(1), 110. e111-110. e117.
- Sforza, C., Grandi, G., Catti, F., Tommasi, D. G., Ugolini, A., & Ferrario, V. F. (2009). Age- and sex-related changes in the soft tissues of the orbital region. *Forensic Science International*, 185(1), 115. e111-115. e118.
- Sforza, C., Grandi, G., De Menezes, M., Tartaglia, G. M., & Ferrario, V. F. (2011). Age- and sex-related changes in the normal human external nose. *Forensic Science International*, 204(1), 205. e201-205. e209.
- Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and Remembering Faces* (pp. 105-131). London: Academic Press.
- Solso, R. L. (2001). Brain activities in a skilled versus a novice artist: an fMRI study. *Leonardo*, 34(1), 31-34.
- Speed, H. (1917). *The Practice and Science of Drawing* (3rd ed.). New York: Dover.
- Starbuck, J. M. (2014). Quantitative Evaluation of the Facial Morphology of a Tolteca Figurine from Mexico using Geometric Morphometric Approaches. *International Journal of Morphology*, 32(2).
- Sturgis, A. (1998). *Faces*. London: National Gallery Publications.
- Třebický, V., Fialová, J., Kleisner, K., & Havlíček, J. (2016). Focal length affects depicted shape and perception of facial images. *PLoS ONE*, 11(2), e0149313.
- von Alphen, E. (1997). The portrait's dispersal: concepts of representation and subjectivity in contemporary portraiture. In J. Woodall (Ed.), *Portraiture: Facing the subject*. Manchester: Manchester University Press.
- Webster, M., & Sheets, H. D. (2010). A practical introduction to landmark-based geometric morphometrics. *Quantitative Methods in Paleobiology*, 16, 168-188.
- West, S. (2004). *Portraiture*. Oxford: Oxford University Press.
- Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science*, 26(3), 212-217.
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737-746.

SUPPLEMENTARY INFORMATION

Production of the 3D print (ACES)

The scanning of our human model involves the use of a hand-held scanning system. The device used produces modulated pulses of light and captures the reflected light, storing the information as a series of points, which can be manipulated in relation to one another. The system, an Artec Eva, allows us to obtain 3D resolution up to 500µm, and 3D accuracy up to 100µm. To obtain an accurate scan of a human face, we must capture facial landmarks on our target, scanning from ear-to-ear across the eye level, and then from ear-to-ear following the jaw-line. The resulting data points have sufficient overlapping facial landmarks to allow the two scans to be stitched together, creating an accurate representation of our model.

SI Figure 1. Fabrication process for 3D printing a 1:1 scale human face

With the digital model captured (SI Figure 1A), we then cropped the scan data to define the required facial features. Setting the boundary line to the base of the jaw, back to the ears, and then along the hair line above the forehead. To successfully print this at a 1:1 scale, the scan had to be split into two components (1B). To ensure a precise assembly, two mounting pins were attached to the back of each ear (1C). Each segment of the face was then thickened to 3mm for ease of printing and structural stability, and then produced using fused deposition modelling. The two parts were then assembled using the pin-holes as guides, the seam filled with putty to conceal the mating face, and subsequently spray-painted grey to enhance contour visibility.

2D Landmark Coordinates and Definitions

SI Figure 2. 2D Landmarks

Homologous landmarks (red) and semi-landmarks (orange) derived from curves (blue connectors).

See SI Table 1 for landmark definitions.

The illustrated image is the $\bar{X}3D$ print.

SI Table 1. 2D Landmark Definitions

Homologous Landmarks		Associated Semi-Landmarks	
No.	Definition	No.	Description
1	Highest point of the forehead in profile	20-22	Lateral left forehead
2	Most lateral point of the brow in profile	23-25	Superior left brow
3	Deepest point of the nose in profile	26-28	Nasal profile
4	Right endocanthion (inner eye corner)	53-55	Upper right eye lid
5	Right exocanthion (outer eye corner)	56	Mid point of lower right eye lid
6	Most lateral point of the nasal tip in profile	29-31	Inferior nasal tip in profile
7	Highest point of the right nostril	-	-
8	Lowest point of the right nasal wing	32-33	Right nasal wing
9	Highest point of the right nasal wing	-	-
10	Intersection of the nose and upper mouth in profile	34-36	Upper mouth in profile
11	Most lateral meeting of the vermillion of the lips	-	-
12	Most inferior point on the lower lip in profile	37	Mid point lower vermillion of lip
13	Right cheilion (outer mouth corner)	-	-
14	Most lateral point of the chin in profile	38-40	Lower mouth in profile
15	Deepest point of the chin in profile	41-43	Lower chin in profile
16	Intersection of interior right earlobe with the outer cheek	44-50	Right lower jaw
17	Lowest point of the earlobe	51	Mid point of lateral lower ear
18	Most lateral point of the right pinna (upper ear)	52	Mid point of lateral upper ear
19	Most superior point of the right pinna	-	-