Perceptions of facial aesthetics in two and three dimensions

S.-A. Todd, P. Hammond, T. Hutton, S. Cochrane and S. Cunningham
Orthodontic Unit, Eastman Dental Institute, University College London, UK

SUMMARY The aims of this study were to investigate whether the preferred facial relationship chosen by professionals and the general public is Class I and to ascertain whether viewing two-dimensional (2D) or three-dimensional (3D) images had any effect on the ranking of facial attractiveness. Orthodontists (n = 47), maxillofacial surgeons (n = 25) and members of the general public (n = 78) assessed 2D and 3D facial scans of two males and two females that had been morphed to produce five images reflecting different skeletal patterns: Class I, mild and moderate Class II, and mild and moderate Class III. Each assessor placed the images in rank order of preference, after viewing alternate 2D and 3D image formats for each face. The data were analysed using logistic regression.

In 2D, professionals (orthodontists and maxillofacial surgeons) chose Class I as the preferred facial image more frequently than the general public for only one of the four faces. However, in 3D format they chose Class I as the preferred facial image for some subject faces more, and others less, frequently when compared with the general public. The gender of the assessor was not significant when assessing the preferred facial relationship for Class I images in either 2D or 3D formats. The oldest assessors (56+ years) were significantly less likely than the younger age groups to select Class I as the preferred facial relationship in both 2D and 3D. In summary, there was too great a degree of variation to say that a difference between 2D and 3D facial images was evident.

Introduction

Previous research using two-dimensional (2D) images has shown that professionals and the general public differ in their perceptions of facial aesthetics, with professionals more likely to prefer a Class I facial profile (Cochrane et al., 1997, 1999). With the increasing use of new technology in medicine and dentistry, three-dimensional (3D) images are now more readily available. However, comparison of 2D and 3D facial image formats has rarely been undertaken.

Facial aesthetics is of great importance to orthodontists, but is a subject that interests and involves people everywhere. An individual’s facial appearance is one of their most obvious characteristics and has profound social significance (Shaw, 1981). Ferrario et al. (1997) believed that the only scientific way of measuring a subjective quality such as facial beauty was through a panel of judges, but they did not state who this panel should include (clinicians, the general public, or both?). The first complicating factor is that the clinician’s perception of ideal aesthetics is not always the same as that of the patient (Tedesco et al., 1983a, b; Cochrane et al., 1997, 1999).

The work of Prahl-Andersen et al. (1979) demonstrated that a significant difference exists between professionals and lay persons when subjectively evaluating morphological characteristics in the dentofacial region. They suggested that the observed differences were probably related to the variation in knowledge and experience between the groups. Similar findings were noted by Kerr and O’Donnell (1990), who found that non-dental professionals were less critical in their appraisal of facial attractiveness compared with dental professionals.

Phillips et al. (1992) undertook a similar study but investigated the effect of different levels of dental training with respect to rating facial attractiveness. The research showed that orthodontic residents rated subjects as being more attractive than other raters. This finding may be due to the specialist training and experience of these orthodontic professionals, allowing extremes of variation to be noted. From this, the authors concluded that the facial attractiveness score may be influenced by the effect of the assessor’s dental training. This was also the conclusion reached by Brismar (1980) when he compared dentists’, dental students’ and patients’ concepts of aesthetics. He found that dentist and patient concepts differed greatly, with those of the dental students being somewhere between the two.

More recent work by Eli et al. (2001) looked at the effect of dental appearance on the development of first impressions by another person. They found that intact and non-decayed dentitions were considered to be more attractive and the subjects were thought to have better aesthetics, social and professional traits when ranked by an assessor of the opposite sex. Another finding from this study was the fact that more women placed appearance as being important when compared with men.

The above studies demonstrate that there is little agreement on the significance of dental and facial appearance. In fact, it allows us to reflect upon and emphasize the point that the appraisal of aesthetic matters is still a very individual concept.
A further issue is how a 3D subject is scrutinized in order to assess dentofacial attractiveness. Moss et al. (1995) noted that millions of pounds are spent on improving or making the most of our faces, but when we view ourselves in the mirror, we usually observe only one aspect, normally a frontal view, of a 3D object.

2D images have been used in the past to assess facial aesthetics, with their obvious limitations in terms of quality of capture and reproduction. Now that non-invasive 3D images are available, clinicians, patients and the general public can assess a more realistic image (Arridge et al., 1985; McCance et al., 1992; Moss et al., 1995).

The aims of this study were to investigate whether the preferred facial relationship chosen by professionals and the general public is Class I, and to determine whether viewing facial images in 2D and 3D has any effect on the ranking of facial attractiveness.

**Subjects and method**

**Sample size**

A sample size calculation, undertaken using nQuery Adviser (Version 4.0, Statistical Solutions, Cork, Ireland) showed that 80 subjects were required in both the professional and general public groups. Data from previous studies assessing facial aesthetics in 2D provided information for the sample size calculation (Cochrane et al., 1997, 1999).

**Stimulus face selection**

Individuals who fulfilled the following criteria were approached and asked if they were willing to participate in the study as ‘stimulus faces’:

1. Over 18 years of age.
2. Caucasian.
3. Skeletal Class II facial dimensions where there was minimal change in the vertical facial dimension when the mandible was postured forward.

Two males (faces A and B) and two females (faces C and D) were selected (age range 28–32 years). Each individual gave consent and underwent a series of black and white facial scans taken using a 3D photogrammetric face scanner (3dMD LLC, Harefield, Middlesex, UK).

**Scanning the individuals**

Each individual had a scan taken in retruded contact position, followed immediately afterwards by another scan in maximum protrusion. The time taken to complete each facial scan was approximately 20 seconds and the processing time was in the order of 2–3 minutes. The scanner captures multiple views that are processed by stereo-matching algorithms to produce a 3D image of a subject.

The images created in retruded and protruded positions for each individual were landmarked with a series of soft tissue points in order to allow the computer software to create an accurate 3D ‘morph’ from the combination of the 2D stereopairs. Shapefind software (Biomedical Informatics, Eastman Dental Institute) was then used to manipulate these images into facial profiles ranging from a moderate Class II to a moderate Class III: moderate Class II skeletal pattern; mild Class II skeletal pattern; Class I skeletal pattern; mild Class III skeletal pattern; moderate Class III skeletal pattern.

A summary of this process is given in Figure 1.

**Assessor selection**

The first research question was concerned with the comparison of groups: professionals and the general public. The professional group was also subdivided to allow
individual comparisons within the orthodontist and maxillofacial surgeon groups.

The general public sample was a convenience sample recruited from the hospital waiting rooms of the Eastman Dental Hospital (excluding orthodontics), local businesses and colleges. The professionals were also convenience samples and were obtained by recruiting at the 2002 British Orthodontic Conference in Glasgow and also travelling to Orthodontic and Maxillofacial Units in the South East of England. The age, gender and number of years experience of the clinicians were recorded in an attempt to recruit a representative sample of that professional group. Financial and ethical constraints precluded truly random recruitment of assessors. Any assessors who knew the individuals presented as stimulus faces were also excluded from participating in the study.

Validation exercise

Before selecting which of the 2D and 3D images to use in the study, they were shown in printed 2D form to 10 senior orthodontic colleagues, in order to establish whether the range of skeletal discrepancies was appropriate. Their feedback from this exercise was that the range of images presented was a suitable representation of each of the skeletal discrepancies.

Main study

In order to avoid introducing bias when showing the stimulus faces to the assessors, alternate male 2D and female 3D images were followed by female 2D and male 3D images. For each assessor, the stimulus faces were presented in a standardized format whereby all five images were shown at once. The five 3D images were presented on a laptop so that they could be rotated and viewed from any desired angle. The five 2D images were presented to the assessor in circular frames to prevent any head posture bias. The assessor was then asked to rank the images in order of preference from 1 (most favourable) to 5 (least favourable). The assessors were given a period of 60 seconds to rank the 2D images and 90 seconds to rank the 3D images. They were not permitted to assign the same ranking to more than one of the profiles.

Statistical analyses

The following statistical analyses were undertaken.

The chi-squared test with continuity correction (Petrie and Watson, 1999).

This was used to investigate the associations between the preferred facial image (in 2D and 3D) and the age, gender and group of the assessor, prior to entering any significant factors in the logistic regression analysis. The level of significance was set at 10 per cent.

Logistic regression analysis (Petrie and Watson, 1999).

Due to the influence of several possible independent variables (gender, age and group), a logistic regression analysis was undertaken to determine whether the assessor group, gender and age were significant factors when rating a Class I profile as the most attractive in both 2D and 3D formats. The level of significance was set at the 5 per cent level. An odds ratio of greater than 1 is a positive finding and less than 1 is negative, with the latter being expressed as a percentage.

Binary variables were allocated values of ‘1’ and ‘0’. When the professional group was subdivided there were more than two categories, so ‘dummy’ variables were required to accurately represent the options. The reference category, against which others were contrasted, was that for which all dummy variables adopted the value 0 and groups were contrasted with the largest group (general public). Similar dummy variables were created for age. The largest age band was 16–35 years, against which the other two age bands of 36–55 years and 56+ years were contrasted. A rotation of dummy variables was undertaken where necessary.

Results

The total number of assessors who took part in the study was 150 (47 orthodontists, 25 maxillofacial surgeons and 78 members of the general public). There was a wide variation in age and gender for each of the groups of assessors. The age groups were 16–35, 36–55 and 56+ years. In both the orthodontist and general public assessor groups the largest age band was the youngest. For the maxillofacial surgeons, the middle age band had the largest number of assessors. In the professional groups there were more male assessors, while in the general public group females dominated. Although it would be ideal to match age ranges and gender for the groups, this was hindered by the difficulty in recruiting the required sample size in each group (particularly the clinician groups).

The four stimulus faces were labelled A–D, whether in 2D or 3D. The two male faces were A and B and the two females faces C and D.

2D images

For the 2D images, the professional group showed significantly increased odds (2.21) of selecting Class I as the preferred facial image, when compared with the general public, but only for face A (Table 1).

When the professional group was subdivided (Table 2), for face A there was again a difference between the assessor groups, with maxillofacial surgeons choosing the Class I image as their preferred image significantly more frequently (odds of 2.68 times greater) than the general public group. However, there was no significant difference between orthodontists and the general public group.
The results for the 3D image formats (Table 1) indicate that the assessor group was significant for faces B and D. For face B, the professionals were less likely (by 61 per cent) to choose Class I than the general public, but their odds of choosing Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.

When the professional group was subdivided (Table 2), for face B both orthodontists and maxillofacial surgeons selected the Class I image as the preferred image significantly less frequently than the general public group. The orthodontists were 59 per cent less likely and the maxillofacial surgeons 66 per cent less likely to choose Class I as the preferred facial image for face D were 2.01 times greater.
was presented with all five images at once. However, ranking the faces in order of preference from 1 (most favoured) to 5 (least favoured) is still subjective and gives no ability to quantify the ‘degree of favourability’ within each group of assessors. However, in agreement with Phillips et al. (1992), it was felt that using this ‘location-free’ rank ordering system to measure facial aesthetics had advantages over other methods, such as visual analogue scales (VAS) (Aitken, 1969; Howells and Shaw, 1985). Although both techniques are convenient and rapid, the actual score recorded on the VAS could be affected by the training and experience of the assessor (Phillips et al., 1992).

Sample size

Consideration must also be given to the sample size in each of the assessor groups, as well as to the number of stimulus faces used in this study. It is acknowledged that there were slightly fewer assessors in both the professional and general public groups than the sample size calculation suggested (72 rather than 80 in the professional group and 78 rather than 80 in the general public group). Subsequently, when the professional group was subdivided, the results should be interpreted with caution, as the sample sizes (especially the maxillofacial surgeon group) were relatively small. However, it was still felt to be useful to analyse the groups independently to establish whether there were any interesting trends. Unfortunately, due to the financial constraints of the study, it was not feasible to travel further and recruit more professionals. As the groups were, therefore, convenience samples, some bias may have been introduced.

In this new area of research, the study was trying to identify trends. For future research, ideally more stimulus faces and more assessors should be used, but this will also increase the time and effort to complete any evaluation and may reduce assessor co-operation.

2D results

Effect of assessor group.

The less critical appraisal of facial aesthetics by the general public (Kerr and O’Donnell, 1990; Cunningham, 1999) was expected to be evident in the findings of this study, i.e. choosing Class I as the preferred facial relationship less frequently than the professionals. The professional group (orthodontists and maxillofacial surgeons) was expected to choose Class I as the preferred facial relationship for 2D images, due to their previous specialty training (Cochrane et al., 1997, 1999). However, the results were not entirely as predicted (Table 1). The professionals had greater odds of choosing Class I as the preferred facial image when compared with the general public, but only for one of the four faces, face A (male).

Of the non-significant findings for the three remaining faces, it is of note that for one, face C (female), professionals were still more likely (odds ratio 1.26) to choose the Class I image when compared with the general public, but this did not reach significance. This highlights the variability of the results, with only one male face being chosen with Class I as the preferred facial relationship by professionals.

It was expected that the three assessor groups, orthodontists, maxillofacial surgeons and the general public, would show significant differences in their chosen facial preferences when compared with each other. This expectation stems from work undertaken by Prahl-Andersen et al. (1979) and Cochrane et al. (1997, 1999). Interestingly, this result was noted for only one of the four faces. In 2D format, face A showed significant findings related to the assessor group (Table 2). After rotation of the dummy variables, the only significant finding was that maxillofacial surgeons had 2.68 greater odds of choosing Class I as the preferred image when compared with the general public. Although there was a non-significant finding for orthodontists when compared with the general public, the statistical results showed that they were still more likely (2.02 greater odds) to choose Class I as the preferred facial image for this particular face.

This partly answers one of the aims of the study by showing that the preferred facial relationship chosen by professionals was different to that of the general public for only one of the four faces. The assessment of facial form is subjective and the variability in the results is perhaps not surprising. Also, caution must be exercised when interpreting the results due to the relatively small sample sizes.

Effect of age of assessor.

Interestingly, the oldest age group of assessors selected the Class I image as the preferred facial form significantly less frequently when compared with the youngest assessors, but only for the female images (faces C and D). This may indicate that the older assessors from all three groups, orthodontists, maxillofacial surgeons and the general public, are more tolerant of individual facial variations in females. It may be that the younger raters have grown up with the concept that alteration of facial aesthetics is an accepted norm, and those persons with displeasing facial features can now electively choose to alter them.

Effect of gender of assessor.

Research carried out by Iliffe (1960) and Udry (1965) looked at facial aesthetics as assessed by British and American assessors, respectively. There were no significant gender differences, and they concluded that no age or occupation differences existed between the two nations. Tedesco et al. (1983b) also failed to find any gender or ethnicity differences between raters and subjects when Black and Caucasian persons were investigated. The results of the present study agree with the findings of Tedesco et al. (1983b), that there were no significant gender differences when assessors chose Class I as the preferred facial relationship.
3D results

Effect of assessor group.

There has been little previous work assessing facial aesthetics in 3D. The results of the 3D component of this study (Table 1) showed that for face D, the professionals chose Class I as the preferred facial image significantly more often than the general public, with increased odds of 2.01. However, the ‘ideal’ Class I facial relationship was chosen significantly less often by both professional groups compared with the general public, in a 3D format for face B. This example highlights the variation within the 3D results for the assessors and the variation between the stimulus faces studied. The logistic regression analysis for individual assessor groups (Table 2) confirmed the above results that, for the male face B, both the orthodontists and maxillofacial surgeons selected Class I as the preferred facial relationship significantly less often than the general public. For face D, when the professional group was subdivided and compared with the general public (Table 2), only the orthodontists were significantly more likely to choose Class I as the preferred image, with odds of 2.37. There were no significant findings for maxillofacial surgeons when compared with either the general public or the orthodontists (Table 2). However, these findings must be treated with caution due to the small sample size.

Effect of age of assessor.

As in 2D, face C (female) proved to be significant for age, with the older age group less likely to choose Class I compared with the youngest age group. This finding may be worthy of further investigation.

Effect of gender of assessor.

The results are similar to those in 2D, in that there were no significant findings for the gender of the assessor when choosing the preferred facial relationship.

Comparison of 2D and 3D results

When the 2D and 3D results were compared, no clear pattern existed. Considering that the analysis was carried out on only four stimulus faces, but by 150 assessors, one may have expected more consistent findings.

Influence of assessor’s background.

One of the original aims of this study was to determine if there was a difference between 2D and 3D image formats when presented to assessors for the evaluation of facial aesthetics. The results indicated that in many instances there was no consistency between the findings for 2D and 3D images or indeed between the four stimulus faces for 2D and 3D formats. An example of this is face A, where the assessor group was significant in 2D but not in 3D. For face B, both professional assessors were significant in 3D, but not in 2D. Although these findings are disparate, it appears that there may be some differences between 2D and 3D images, in the assessors’ choice of Class I as the preferred facial relationship, and these findings need to be investigated further with larger sample sizes.

Age of assessor.

Another partially consistent finding between 2D and 3D images was that of the age of the assessor. The oldest age group of assessors (56–77 years) was less likely to choose Class I as the preferred image for female faces C and D in 2D and face C in 3D. A further investigation would be required, with an increase in both subject and assessor sample sizes, to draw meaningful conclusions from these initial findings.

Gender of assessor.

Some findings from this study indicate that there was continuity when selecting the preferred facial relationship from both 2D and 3D image formats. Throughout all the analyses, in both 2D and 3D, there were no significant findings for the gender of the assessor when choosing the preferred facial relationship.

As orthodontists, we may apply our own source of aesthetic values to the exclusion of those of our patients. Orthodontists do, and should, play a deciding role in determining the aesthetic destiny of a patient’s face, but they must take into account the patient’s perception of their own face before treatment planning. In this way, orthodontists are obliged to study facial beauty, harmony, balance, and proportion as perceived, not just through their own eyes, but also through those of the general public.

However, there are multiple 2D, and now 3D, image formats for raters to assess facial and dental attractiveness. The question that arises is whether one of these methods is better than any other. There still appears to be a large variation in both professional and lay persons’ attitudes to dental and facial attractiveness, irrespective of the format in which the image is presented. This serves to inform us that the nature of beauty and facial attractiveness is still a very complex and subjective measurement, requiring future research in both the psychosocial and clinical perspectives.

Conclusions

The results indicated that there was no consistency between the findings for 2D and 3D images or indeed between the four stimulus faces for 2D and 3D.

For some images, professionals (orthodontists and maxillofacial surgeons) were more likely to choose Class I as the preferred facial image in 2D when compared with the general public, but it must be stressed that this was not for...
all the faces. In 3D format, they chose some subject faces more, and others less, frequently when compared with the general public for Class I as the preferred facial form.

The oldest assessors (56+ years) were less likely to select Class I as the preferred facial relationship in 2D and in 3D for the female faces.

The gender of the assessor was not significant when assessing the preferred facial relationship for either Class I or mild Class II images in either 2D or 3D formats.

Address for correspondence
Susan Cunningham
Department of Orthodontics
Eastman Dental Institute
256 Gray’s Inn Road
London WC1X 8LD
UK
E-mail: s.cunningham@eastman.ucl.ac.uk

Acknowledgement
The authors would like to thank Dr Aviva Petrie for her statistical advice.

References
Cunningham S J 1999 The psychology of facial appearance. Dental Update 26: 438–443