ABSTRACT

Forty-three women and 45 men rated a series of images of selected paintings and sculptures for aesthetic appeal and originality. For each painting or sculpture, there were three versions: the original image and two manipulated images, one displaying a lower waist-to-hip ratio (WHR) and the other displaying a higher WHR than in the original. The results showed that, in general, both male and female participants considered the original and, in some cases, the image with the higher WHR as the most aesthetically pleasing. The results also showed that participants were generally able to discern the unaltered image as being the original, although women were better at this task than men. Implications for the study of aesthetics and physical attractiveness are considered.

INTRODUCTION

A fundamental question in aesthetics concerns the properties and qualities of a system, image, or object that appeal to the senses. That is, can there be a science of
aesthetics that is able to specify the attributes of a face, body, or object that is most beautiful or pleasant? Historically, various attempts have been made to answer this question (Armstrong, 2005; Eco, 2005; Swami, 2007). Psychological studies of an experimental nature can be traced back to pioneering investigations on the “golden ratio” by Gustav Fechner (1878). More recent studies in this tradition have examined preferences for very simple stimuli such as rectangles and triangles (McManus, 1980) and color (McManus, Jones, & Cottrell, 1981).

Other studies have looked at the balance and composition of paintings, which concerns the arrangement and inter-relation of elements, rather than with the elements themselves (e.g., Locher, 2003; Locher, Gray, & Nodine, 1996; Locher, Stappers, & Overbeeke, 1999; McManus, Edmondson, & Rodger, 1985; Nodine, Locher, & Krupinski, 1993; Vartanian, Martindale, Podsadlo, Overbay, & Borkum, 2005). For instance, McManus (1987) asked participants to make preference judgments for simple compositions in which a single object (a boat or a church) was placed at different positions within a frame. Locher, Smith, and Smith (1999, 2001) showed that aesthetic judgements of paintings are a function partly of presentation format (e.g., gallery, slide, and computer generated images).

The assumption behind these studies is the idea that there is a visual ideal associated with every composition, and that deviations from that ideal are noticeable (Arnheim, 1982; Locher, Stappers et al., 1999). One way to test whether a work of art is in that ideal state is to compare it with alternative compositions (Kulka, 1989). If the alternative composition is judged to be better than the original, than it might be suggested that the alternative more closely approximates the ideal than the original from which it was derived. A number of studies have attempted to test this idea by subjecting works of art to experimental analysis.

For instance, McManus, Cheema, and Stoker (1993) modified 25 paintings by Piet Mondrian, and found that participants could reliably distinguish the originals from the pseudo-Mondrian (see also Noll, 1966). A recent replication showed that the Mondrian results are robust, although participants had greater difficulty differentiating original and manipulated paintings by Damien Hirst (Furnham & Rao, 2002). Finally, Vartanian, Overbay, Podsadlo, and Martindale (2000) pitted 15 works of art by van Gogh, Matisse, Malevich, and Pissarro against compositionally-altered counterparts, and in each case, the originals were preferred over the alternatives.

These studies provided the theoretical background for the present investigation. However, rather than manipulate the composition of famous paintings, we manipulated a single attribute that has been the subject of much debate within the psychological literature, namely the waist-to-hip ratio (WHR). Some evolutionary psychologists have proposed that the WHR of human beings evolved under sexual selection, and that contemporary men have ‘psychological modules’ that elicit a preference for a low WHR in women (Singh, 1993; for reviews, see Swami & Furnham, 2006, 2007).
The preference for a low WHR has been replicated in a wide range of cultures using different methodological tools, but has also generated critical debate. The latter has centered around the relative contributions of the WHR and overall body weight to judgments of physical attractiveness, the evolutionary significance of a preference for low WHRs and alternative, non-evolutionary explanations for both the preference for low WHRs and certain body weights (see Swami, 2007; Swami & Furnham, 2007).

Notwithstanding these methodological and experimental issues, the present study manipulated the WHRs of five paintings and sculptures respectively to examine its effect on aesthetic appeal. There is reason to believe that this manipulation should be meaningful. For example, it has been suggested that, despite local aesthetic canons, artists cross-culturally and cross-generationally have represented female WHRs in the healthy and fertile range (Singh, Davis, & Randall, 2000). That is, despite differences in local taste, cultural conditions, and economic factors, almost all cultures and generations share a particular notion a low WHR (typically centered around .70) as healthy and, therefore, attractive.

Thus, Singh et al. (2000) measured the WHRs of some 300 ancient sculptures from India, Egypt, Greece, and some African tribes, and found a high degree of similarity in depictions of the WHR in all four cultural groups. More recently, however, Swami, Gray, and Furnham (2007) measured the WHRs of 29 women in paintings by Pieter Pauwel Rubens, and found that the mean was significantly higher than the value of .70, which is usually taken as being “optimally attractive.” This suggests that the preference for a low WHR may not be as universal as Singh (1993) has claimed, although more work is needed in this regard.

The present study was designed to extend previous research. The images of five paintings and sculptures respectively were digitally manipulated to yield a lower or a higher WHR. Based on previous studies, we formulated two competing hypotheses. First (H1), the images in the present study with the lowest WHRs would be considered most aesthetically-pleasing. In a previous study which made similar manipulations to images of real women, Henss (2000) found that the image with the lowest WHR was considered most attractive, fertile, youthful, and healthy. By contrast (H2), if the images used in the present study reflect some “optimal aesthetic configuration” (Locher, 2003; McManus et al., 1993), then we might predict that the original images will be considered most aesthetically pleasant.

**METHOD**

**Stimuli**

The stimuli were created from color images of five paintings and five sculptures, all of which were downloaded from the Internet. Prior to the study, a wide range of images of paintings and sculptures were selected by a knowledgeable art historian for preliminary manipulation. Only images in which the frontal WHR
was clearly visible were selected for use by the experimenters in the final sample. The final sample consisted of the five paintings and sculptures, respectively, which allowed for ease of manipulation of the WHR, without compromising the overall composition, style, or elements of the image. The final five paintings selected were William-Adolphe Bouguereau’s *The Bather* (1870) and *The Birth of Venus* (1879), Paul Gaugin’s *Annah the Javanese* (1893), Ambrose Duvel’s *The Birth of Venus* (1862), and Jean Auguste Dominique Ingres’ *La Source* (1856). The final five sculptures were three ancient Greek statues, namely the *Vénus de Milo*, the *Vénus of Arles* and the *Esquilin Vénus*, and Auguste Rodin’s *Toilette of Venus* (1885) and *Venus* (1883).

For each image, two versions of the WHR were created using Adobe’s Photoshop, a commercial software widely used to produce photo-real imagery which is indistinguishable from what would be termed a real photograph. In one picture, the WHR was tightened and in the other the WHR was widened. This created three images for each painting or sculpture: the original image and two manipulated images (one displaying a lower WHR and the other displaying a higher WHR). Example of the images used in this study is provided in Figure 1.

It should be noted that the amount of manipulation of the WHR was slightly different for the different stimulus images. To measure the WHRs, we followed Henss (2000) in enlarging each image and measuring the smallest hip and broadest waist widths. It should be noted that this measure of the WHR takes the distance across the waist and divides it by the distance across the hips. The measure that has been correlated with female fertility in some studies, on the other hand, measures the distance around the waist divided by the distance around the hips. This is an acknowledged limitation of two-dimensional stimuli (Henss, 2000), but it should also be pointed out that most psychological studies of the WHR have also made use of WHR_across rather than WHR_around, particularly those using line-drawn stimuli. The WHRs of the images in this study are listed in Table 1.

**Participants**

The participants of this study were 43 women and 45 men ranging in age from 18 to 54 (*M* = 28.99, *SD* = 8.01). The participants were recruited opportunistically by the authors of this study. The mean body mass index (BMI, a measure of body weight) of the male participants was 23.71 kg/m\(^2\) (*SD* = 3.64) and of the female participants was 22.38 kg/m\(^2\) (*SD* = 3.30). The majority of the sample was of British or European Caucasian descent (78.7%), with a small

---

1 In alterations of this nature, it is difficult to alter the WHR of each image to the same degree (see also Henss, 2000). The primary reason for this is that the original WHR of figures may differ from each other. Attempting to systematically vary the WHRs of each figure (e.g., by 5% tightening or widening) may have resulted in differing degrees of alteration to the composition of a painting or sculpture. In the present study, we were keen to limit any change to the overall composition, while also ensuring that the WHR alterations were meaningful.
Figure 1. Examples of the stimuli used in this study. This set represents the original (center) and two manipulated versions of Bouguereau’s *The Barber* (low WHR left, high WHR right).
minority of participants of Asian (13.5%), Afro-Caribbean (1.1%) or other
descent. Participant age, BMI, or ethnicity did not have a significant effect in
subsequent analyses. All participants took part on a voluntary basis and were not
remunerated for their time. Participants were naïve to the aims of the study and
were debriefed following the experiment.

Procedure

All participants were tested individually and were asked to rate each image
presented on a computer screen for 15 seconds. In a pilot study ($n = 8$), the length
of presentation of each image was varied in order to ensure enough time for
participants to make their ratings without making the manipulation of WHR
obvious. A presentation time of 15 seconds was found to balance both needs
optimally. In the actual experiment, the order of presentation of each image was
randomized completely so that no image of the same painting or sculpture was
presented sequentially. That is, all 30 images were randomly presented, and the
three versions of each painting or sculpture did not appear consequentially.

Participants were presented with a questionnaire on which they made their
ratings. They were told that each image was different from every other, and that
some images had been doctored while other represented the original painting
or sculpture. For each image, participants were asked to provide two ratings:
1) how aesthetically pleasing (beautiful) they believed the image to be; and
2) how likely they believed the image depicted to be the original painting
or sculpture. Ratings were made on a 7-point Likert scale (1 = Aesthetically

<table>
<thead>
<tr>
<th>Image</th>
<th>Smaller</th>
<th>Original</th>
<th>Wider</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>The Bather</em> (Bouguereau)</td>
<td>.65</td>
<td>.70</td>
<td>.80</td>
</tr>
<tr>
<td><em>The Birth of Venus</em> (Bouguereau)</td>
<td>.60</td>
<td>.70</td>
<td>.82</td>
</tr>
<tr>
<td><em>Annah the Javanese</em> (Gaugin)</td>
<td>.83</td>
<td>.90</td>
<td>.96</td>
</tr>
<tr>
<td><em>The Birth of Venus</em> (Duvel)</td>
<td>.67</td>
<td>.76</td>
<td>.80</td>
</tr>
<tr>
<td><em>La Source</em> (Ingres)</td>
<td>.65</td>
<td>.75</td>
<td>.81</td>
</tr>
<tr>
<td><em>Vénus de Milo</em></td>
<td>.67</td>
<td>.71</td>
<td>.77</td>
</tr>
<tr>
<td><em>Vénus of Arles</em></td>
<td>.73</td>
<td>.83</td>
<td>.90</td>
</tr>
<tr>
<td><em>Esquilin Vênus</em></td>
<td>.73</td>
<td>.79</td>
<td>.84</td>
</tr>
<tr>
<td><em>Venus</em> (Rodin)</td>
<td>.67</td>
<td>.75</td>
<td>.79</td>
</tr>
<tr>
<td><em>Toilette de Venus</em> (Rodin)</td>
<td>.71</td>
<td>.76</td>
<td>.80</td>
</tr>
</tbody>
</table>
unpleasant/Doctored image, 7 = Aesthetically pleasing/Original image). The testing session lasted around 15 minutes.

RESULTS

Cross-Ratings Correlations

As a simple test of whether the ratings of aesthetic appeal and originality were correlated, we measured Pearson’s correlations for each set of ratings for each image. The results showed no significant correlations ($p > .05$) for any of the pairs of ratings. This remained true when the WHR ratings were collapsed across images ($p > .05$). It appeared, therefore, that participants were making the two ratings relatively independently of one another.

Analysis of Variance

For both sets of ratings, a $10 \times 3 \times 2$ mixed model analysis of variance (ANOVA) was computed. Image and WHR were treated as within subjects factors and participant gender was treated as a between subjects factor. Where a factor had more than two levels, and the Mauchly’s Test of Sphericity was shown to be significant, the Greenhouse-Geisser correction was applied. Summaries of the ANOVA results and the main effects of Image, WHR, and Gender and their interactions for aesthetic appeal and originality are shown in Tables 2 and 3 respectively.

Aesthetic Appeal

The ANOVA for aesthetical appeal showed that both Image and WHR had a significant effect on the ratings of stimuli. Partial eta squared values revealed that

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>$F$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>6.67, 573.85$^a$</td>
<td>22.30**</td>
<td>.21</td>
</tr>
<tr>
<td>Image \times Gender</td>
<td>6.67, 573.85$^a$</td>
<td>1.05</td>
<td>.01</td>
</tr>
<tr>
<td>WHR</td>
<td>1.86, 159.55$^a$</td>
<td>7.65*</td>
<td>.08</td>
</tr>
<tr>
<td>WHR \times Gender</td>
<td>1.86, 159.55$^a$</td>
<td>1.45</td>
<td>.02</td>
</tr>
<tr>
<td>Image \times WHR</td>
<td>9.86, 847.71$^a$</td>
<td>10.62**</td>
<td>.11</td>
</tr>
<tr>
<td>Image \times WHR \times Gender</td>
<td>9.86, 847.71$^a$</td>
<td>1.30</td>
<td>.02</td>
</tr>
</tbody>
</table>

$^a$ Greenhouse-Geisser corrected; *$p < .05$, **$p < .001$. 

Table 2. ANOVA Results with the Main Effects of Image, WHR, and Their Interactions for Ratings of Aesthetic Appeal
Image accounted for the greater amount of variance in the data ($\eta^2_p = .21$) when compared with WHR ($\eta^2_p = .08$). This suggests that the actual Image was more important than WHR when judging aesthetic appeal, and suggests that some paintings/sculptures were rated as more pleasant than others. There was also a significant interaction of Image and WHR ($\eta^2_p = .11$). However, the between subjects results showed that gender did not have a significant interaction with any of the within subjects factors. These results are presented in Figure 2.

Considering the Images individually, it can be seen from Figure 2 that there was considerable variability in participants’ ratings. The highest ratings were provided for two of Bouguereau’s paintings, *The Bather* and Bouguereau’s *The Birth of Venus*, and the lowest overall ratings were given to Gaugin’s *Annah the Javanese*. In general, there did not seem to be any overall pattern of ratings in terms of paintings or sculptures, with both types of images being rated relatively similarly. In terms of WHR, the data is more complicated but do not appear to support $H_1$, that is, there is no consistent preference for low WHRs. The lowest WHR received the highest rating for only one image (*Annah the Javanese*). In contrast, the original WHR was rated most positively for six of the images (*The Bather*, Bouguereau’s *The Birth of Venus*, *La Source*, the *Vénus de Milo*, the *Vénus of Arles* and the *Esquilin Vénus*), which lends some support for $H_2$. Somewhat surprisingly, the highest WHR was given highest rating for three of the images (Duvel’s *The Birth of Venus* and the two Rodin sculptures).

Explaining the Image $\times$ WHR interaction is clearly more complicated. For some images, there was a clear pattern of the original WHR being given the most positive rating, followed by the low WHR and finally the high WHR. However, for other figures, this pattern of findings was reversed. In general, then, it can be said that the actual image matters when explaining the effect of the WHR manipulation. This may not seem particularly surprising, for reasons discussed below.

### Table 3. ANOVA Results with the Main Effects of Image, WHR, and Their Interactions for Originality Rating

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Partial eta squared ($\eta^2_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>6.79, 583.47&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.23**</td>
<td>.14</td>
</tr>
<tr>
<td>Image $\times$ Gender</td>
<td>6.79, 583.47&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.77*</td>
<td>.04</td>
</tr>
<tr>
<td>WHR</td>
<td>1.93, 166.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>24.49**</td>
<td>.22</td>
</tr>
<tr>
<td>WHR $\times$ Gender</td>
<td>1.93, 166.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.81*</td>
<td>.09</td>
</tr>
<tr>
<td>Image $\times$ WHR</td>
<td>12.28, 1055.78&lt;sup&gt;d&lt;/sup&gt;</td>
<td>6.72**</td>
<td>.07</td>
</tr>
<tr>
<td>Image $\times$ WHR $\times$ Gender</td>
<td>12.28, 1055.78&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5.02*</td>
<td>.06</td>
</tr>
</tbody>
</table>

<sup>a</sup>Greenhouse-Geisser corrected; *p < .05, **p < .001.
Figure 2. Mean ratings for each image for Aesthetic Appeal.

Note: Bou = Bouguereau, Duv = Duvel; WHR 1 = Smaller, 2 = Original, 3 = Wider.
Originality

As before, the ANOVA for originality showed that Image and WHR had significant main effects. In contrast to the results for aesthetic appeal, however, WHR was found to have a larger partial eta squared value than Image when rating originality (Image $\eta_{p}^2 = .14$; WHR $\eta_{p}^2 = .22$). This suggests that the manipulation of the WHR may have been important in determining the originality of a painting or sculpture. There was also a significant Image × WHR interaction, although the partial eta squared value of this result was relatively small ($\eta_{p}^2 = .07$). Finally, the between subjects results showed that there was a significant interaction of gender with both factors. These results are depicted in Figures 3a and 3b.

For originality, the results for Image showed no discernable pattern of results. Male participants gave the highest ratings for Bouguereau’s *Birth of Venus* and the *Vénus de Milo*. For female participants, the highest overall ratings were given to Bouguereau’s *The Bather* and the *Esquilin Vénus*, although several other images were rated quite highly. In terms of WHR, men correctly rated the undocorated WHR as original for three images (*La Source*, *Vénus de Milo*, and the *Esquilin Venus*). By contrast women correctly identified the original image as being the undocorated version for 7 of the 10 images. This is an interesting finding, and suggests that women may be better at discerning original, undocorated paintings and sculptures. In general, the lowest WHR was considered unoriginal by both male and female participants. There was also an Image × WHR interaction, although this result showed no clear pattern: for some images, the original WHR was given the highest rating, whereas for others the higher WHR was rated most positively.

DISCUSSION

At the outset of this study, we formulated two hypotheses, only one of which was partially supported. First (H1), our results showed that manipulating the WHRs of famous paintings and sculptures did not lead to the lowest WHR being judged as most aesthetically pleasing. Rather, the results showed that the lowest WHR was only rated the most aesthetically pleasant for one painting (*Annah the Javanese*). By contrast, the original or undocorated WHR was considered most aesthetically pleasant for six of the images by both women and men alike, which seems to support H2. It is possible, therefore, that there is an “optimal aesthetic configuration” (McManus et al., 1993) when it comes to the human body in art, and painters or sculptures seem to be able to perceive this.

At first glance, the finding that the lowest WHR was not considered most aesthetically pleasant for the majority of the images may seem surprising. After all, a previous study using a similar method of manipulation, but with human figures, showed that the smallest WHR was consistently judged as the most physically attractive (Henss, 2000). Nevertheless, it should be borne in mind that the ratings for the different images in the present study were relatively
homogenous; that is, with few exceptions (Vénus de Milo and Vénus of Arles in particular), the ratings for the images with smaller, original and wider WHRs were relatively similar. Moreover, the significant WHR × Image interaction would seem to suggest that the manipulation of the WHR mattered, but only when considered in conjunction with the actual image. And this manipulation was differentially effective for the different images in question, although the hypothesized preference for smaller WHRs was not supported.

The results also showed that judgments about originality did not seem to be associated with judgments of aesthetic appeal. That is, participants appeared to be making judgments about the originality of an image independently about their beliefs about the aesthetical appeal of an image. However, images with original (undoctored) and wider WHRs seemed to be judged as original for most of the paintings by men and women, although women appeared to be better suited to this task. This result is particularly dramatic when considering some of the sculptures used as stimuli in the present study (in particular the Vénus de Milo, the original of which was considered highly original by both women and men).

One explanation for this is that the paintings and sculptures used in the present study were relatively well-known, and so participants were able (either consciously or otherwise) to discern the original image. Thus, to the extent that images of the Vénus de Milo are widely available and ubiquitous, we might expect that participants are knowledgeable about its relative proportions. This may also explain why female participants were better able at discerning the original from the manipulated images. If there are gender differences in art appreciation (cf. Perry, 1999), then it might be expected that the more art-appreciating gender will be better able at discerning original images. In future studies, it may be useful to ask participants to also indicate how well they know a painting or sculpture, as this would provide a direct measure of the impact of familiarity on aesthetic judgments.

It seems unlikely that the results of this study were caused by the manipulation being apparent to participants. Indeed, every effort was made to ensure that the manipulation was subtle, and the fact that the wider WHR was considered original for some sets of paintings and sculptures suggests that this manipulation was successful. It is possible that the slightly different manipulation of WHR accounts for the overall results seen in the study (see Table 1). For example, it is possible that the lower WHR manipulation altered the WHR in such a way that it fell below a certain threshold, so that it was no longer considered appealing. Indeed, Furnham and Radley (1989) have pointed out that the relationship between attractiveness and WHR is not strictly monotone in real life. Beyond a certain point, an extremely low WHR may appear grotesque and repelling.

However, given that the manipulation of smaller WHRs in this study showed a good deal of variability, this seems an insufficient explanation of the results. Moreover, it might be expected that the higher WHR should show the same drop-off in aesthetical appeal, but this was the case. Certainly, we acknowledge
Figure 3(a). Mean ratings for each image for Originality (male participants).

Note: Bou = Bougeaureau, Duv = Duvel.
Figure 3(b). Mean ratings for each image for Originality (female participants).

Note: Bou = Bouguereau, Duv = Duvel; WHR 1 = Smaller, 2 = Original, 3 = Wider.
the limitation of somewhat arbitrarily manipulating the WHRs of the images, but we are fairly confident that this alone does not account for the findings of this study. Rather, in general terms, the findings seem to suggest that the original images (and in some case, images with a higher WHR) were considered both aesthetically pleasant and original by participants.

The methodology adopted in this study provides a unique way of approaching the question of aesthetics, but leaves open the question of what any “optimal configuration” might consist of. The most likely suggestion is that the WHR of figures is merely one component among many that define a work of art (see McManus et al., 1993, for variants). For example, Swami (2007) has considered that, when looking at a work of art, we do not tend to break down the image into its individual components. Rather, we notice the whole image and everything that entails. The search for beauty, Swami (2007) proposes, should not be driven by the search for beauty in isolation, in the hope that one quality (such as the WHR) will provide an insight into the meaning of an image (or for that matter, an individual). Rather, the many qualities that define an image are what matter, and more so, they only matter when considered as a whole and not in isolation. But this “law of the whole” (Armstrong, 2005) also suggests that we are more than the sum of our parts: it is the integration and not the summing of parts that matters.

Almost every work of art encapsulates such “wholes” and it should be clear that the different elements work together to create the effect that they do. More practically, the results of this study have an important message for researchers studying the physical attractiveness of human beings. While it may be methodologically easier to examine attributes such as the WHR in isolation, ultimately physical attractiveness—like the appeal of great works of art—can only be understood in its entirety.

REFERENCES


Direct reprint requests to:

Dr. Viren Swami  
Dept. of Psychology  
University of Westminster  
309 Regent Street  
London W1B 2UW  
e-mail: virenswami@hotmail.com