Importance of Genetic Effects for Characteristics of the Human Iris

Mats Larsson1, Nancy L. Pedersen2, and Håkan Stattin1

1 Department of Behavioral, Social and Legal Sciences, Center for Developmental Research, Örebro University, Sweden
2 Department of Medical Epidemiology, Karolinska Institutet, Stockholm, Sweden

The relative importance of genetic influences (heritability) on five general textural quality characteristics of the human iris was assessed using sex and age limitation models. Colour photographs of irises were available from 100 monozygotic twin pairs, 99 dizygotic twin pairs, and 99 unrelated randomly paired age-matched German subjects. Comparative scales were constructed and two judges who were blind to zygosity independently rated five characteristic of the subjects’ left iris. Inter-rater reliabilities were larger than .90 for all five scales. The heritabilities for the five iris characteristics ranged from .51–.90. No sex-specific genetic factors were found for the iris characteristics. Age-group differences in heritability were found for one of the five iris characteristics — “distinction of white dot rings”. Heritability was greater for the older cohort (90%) than the younger (73%). The iris characteristics that showed the next highest additive-genetic effect were “contractional furrows” (78%) and “frequency of crypts” in the main stroma leaf (86%).

Past results are suggestive but before any conclusions concerning the validity of the relationships can be made, the relative importance of genetic effects for different textural quality characteristics in the iris must be demonstrated. The present study investigates whether textural differences in the iris can be used as a biomarker, presumably for a cluster of genes that is expressed in the iris (and putatively in the CNS), by estimating the relative importance of genetic effects for five general textural characteristics. We evaluated the nature of individual differences in five iris features in a sample of like and unlike-sexed twins and unrelated randomly paired age matched dyads.

Materials and Methods

Sample

Data for this study came from a German sample (Burkhardt, 1992) of 100 monozygotic twin pairs (54 male-male, 46 female-female), 99 dizygotic pairs (27 male-male, 34 female-female, 39 male-female) and 99 age-matched randomly paired, unrelated subjects (22 male-male, 28 female-female, 49 male-female). The mean age for the MZ, DZ and unrelated subjects was 20.1 (5–70 years, SD = 12.3); 20.5 (4–72 years, SD = 12.3) and 27.5, (4–78 years, SD = 9.2), respectively. All were healthy volunteers who were recruited through advertising in the town Braunschweig close to Hanover, Germany. The human subjects committee at Braunschweig University reviewed the research protocol. Black and white photographs were taken of the face (front and profile), mouth, eye area, nose and ears, and close-up color photographs were taken of the subjects’ irises.

Zygosity was established by comparing intra-pair similarity for 10 physical characteristics of the head, including eye color, hair type, and shape of the ears. The attributes were selected on the basis of the diagnostic rules developed by Nichols & Bilbro (1966). Two of the twin pairs who claimed to be monozygotic had too large intra-pair differences to be accepted by the algorithm and were excluded from the sample.

Address for correspondence: Mats Larsson, Department of Behavioral, Social and Legal Sciences, Center for Developmental Research, Örebro University, 701 82 Örebro, Sweden. Email: mats.larsson@bsr.oru.se
Material

A stereomicroscope (ZISS-Universalsplltlampe 30 SL/M) with an attached camera (Pentax 1000) with a 125 mm lens was used to obtain color photographs of the subjects’ irises. The exposure time was < 1/1000 seconds and set automatically by the flash exposure control. The shutter speed on the camera was set to 1/60 and the aperture was set to 32. All rolls of film used for the twins had the same charge number. Close-up color photos (where the diameter of the iris on the slide was about 22 mm) were taken of both irises from all subjects. The photo slides of the subjects’ irises were placed on 12 CD-rom discs. The quality of the transformation from slide positives to the images on the CD-rom was chosen through Kodak’s Photo-CD (5 levels, 3072 × 2048 pixels). The photos were viewed on a high contrast color computer screen (Brand: Eizo; Model: FlexScan F55) with 1024 × 768 / 85 Hz resolution (0.28 mm Dot Pitch CRT; fH:27-70 kHz/fV: 50-120 Hz) using the software program Photo Shop 4.0. Iris photographs and other characteristics of the sample were purchased from Dr Angelica Burkhardt, Institute for Human Biology, Technical University Caroli-Wilhelmina in Brunswick, Germany.

Comparative Scale Construction

For each iris characteristic, selected photographs were chosen from the entire sample of all pictures to represent the endpoints of each comparative scale. The photos representing the endpoints were found by looking through all photos in the sample and pinpointing the most representative among the most extreme on the iris characteristic of interest. Photos in the steps between the endpoints were also selected to represent points distributed between endpoints, as equally as possible. Five continuous scales were constructed in this way by the first author. All scales had five scale steps except the scale for eye color which had four scale steps (blue, grey, hazel and brown). The iris characteristics of interest for each scale are illustrated in Figure 1.

The Rating Catalog

Specially trained raters independently reviewed the photographs of the left iris and judged which scale step to which each photo was most similar. The raters judged the number, size, and distinctiveness of the crypts, as well as the nevi, the white dots, the contractional furrows, and the different shades of color. Two raters, blind to zygosity, reviewed each photograph. In order to secure reliable estimates, detailed rules were created for the raters. The rules made reference to photos that exemplified what the raters should do when they were confronted with iris photos that lay between two scale steps. From 8 to 13 photos per scale were selectively chosen for this purpose. The written decision rules and the photos for the scales are available from the first author on request. In all, seven photos from MZ twins, eight photos from DZ twins, and 75 photos from the group of unrelated individuals were used as examples in the scales or as reference photos. These photos were included in the twin analyses but not used in reliability testing.

Rating Procedure

At all times, the ratings were performed in a room with subdued soft lighting. The raters sat approximately 0.50 meter from the computer screen. All photos were presented at the same magnification (i.e., 66.7% — the diameter of the iris spots were 120–130 mm), except for the scale that measured nevi. The magnification for all photos was 100% (i.e., the diameter of the iris spot was then 180–190 mm) for ratings of nevi. The scale was presented to the left of the screen. The photos that exemplified what the raters should do when they were confronted with iris photos that lay between two scale steps were accessible as bookmarks presented at the bottom of the screen. The photos to be rated were randomly sorted into catalogs, and 100 were rated in

Figure 1

The numbered arrows point toward the iris characteristics of interest. 1 crypts in the main stroma; 2 nevi; 3 iris color; 4 a white dot ring, and 5 a contractional furrow.
each rating session. After each session, the reliability was checked and discrepancies between the ratings were discussed together with the test leader. Inter-rater reliability for the scales ranged from .91–.97.

Analysis

The raters’ judgment of which scale step each photo was most similar to was scored as a continuous variable. All five scales were positively skewed, and for this reason the scores were log-transformed. A regression procedure was then applied to the transformed scores to produce residuals controlling for sex and age, respectively (McGue & Bouchard, 1984). These residuals were thereafter used to compute variance-covariance matrices for MZ, like-sexed DZ, and unlike-sexed DZ pairs, using SPSS.

To use all twin groups simultaneously, estimate genetic and environmental parameters, and test for sex specificity and age effects, maximum-likelihood model fitting using the structural equation modeling package Mx (Neale, 1999) was undertaken. The variance-covariance matrices used for testing sex specificity were based on the residuals controlling for age and the matrices for testing age effects were based on residuals controlling for sex.

In order to (1) estimate the magnitude of genetic and environmental effects in males and females separately and (2) determine whether the same set of genes or shared environmental influences characterize the males and females, four models were tested (Neale, 1999). For variables where non-additive genetic effects were indicated (i.e., \( r_{DZ} < \frac{1}{2} r_{MZ} \)), the models included additive genetic factors \( (a^2) \), non-additive genetic factors \( (d^2) \), and non-shared environmental factors \( (c^2) \). For the other variables (i.e., \( r_{DZ} > \frac{1}{2} r_{MZ} \)), shared environmental factors \( (c^2) \) were included, but not non-additive genetic factors \( (d^2) \).

The first model allowed different values for the three parameters for males and females and also estimated a male specific genetic effect \( (a'_m, d'_m, c'_m) \). This model allowed the estimates of the relative importance of the three latent factors (i.e., values of \( a', c', d' \)) to differ for men and women, as well as a male specific genetic effect indicating qualitative differences. The second model was like the first except that the male specific parameter was fixed at 0, testing whether or not the qualitative difference was significant. In the third model parameters \( a', c', d' \), and \( e' \) were fixed to be equal for men and women, and \( a'_m \) or \( d'_m \) fixed to 0. The fourth model was like the third except that the shared-environmental or non-additive genetic parameter for men and women was fixed at 0 — the most parsimonious model.

In order to test for age effects, the twin material was divided into two age groups where the younger cohort was between 4–24 years \( (n_{MZ} = 74; n_{DZ} = 78 \) pairs), and the older was between 25–72 years of age \( (n_{MZ} = 26; n_{DZ} = 21 \) pairs). Three models were used to test for potential age effects (Neale & Cardon, 1992). The first model estimated the parameters \( a^2 \) and \( e^2 \) using the information from 4 twin groups simultaneously (both the younger and older cohorts of MZ and all DZ twin pairs) constraining the estimates to be the same. The second and the third model estimated the parameters \( a^2 \) and \( e^2 \) for the older and younger cohort, respectively. A formal test of heterogeneity is then \( \chi^2_{model 1} - (\chi^2_{model 2} + \chi^2_{model 3}) \). Testing for age effects was only performed for those iris characteristics for which no sex-specific factors were found.

Results

Inter-rater Reliability

Two raters judged each photograph for frequency of crypts in the iris stroma leaf, iris color, distinction of white dot rings, and distinction of contractional furrows in the iris. For the scale that measured frequency of nevi, the inter-rater reliability for the first 51% of the sample was so high (.97) that it was evident that it would be sufficient to have only one judge rate the rest of the sample.

The inter-rater reliabilities for scales 1 to 5 are shown in Table 1. The Spearman-Brown inter-rater reliability, which estimates the reliability of the average individual rater (mean reliability), was consistently high for all five scales. The Spearman-Brown aggregated reliabilities, which estimate the reliability of two raters simultaneously, were also very high.

Descriptive Statistics

Table 2 summarizes means and standard deviations of the five iris characteristics for the MZ and DZ twin pairs. Mean differences between MZ and DZ twins were found on the scale that measured frequency of nevi and distinction of contractional furrows. Likewise, variances differed on the scale that measures iris color. No other mean or

---

### Table 1

Inter-rater Reliability for Scales of Iris Characteristics

<table>
<thead>
<tr>
<th>Scale/Iris Characteristics</th>
<th>Number of Raters/Observations</th>
<th>Inter-rater Reliability¹</th>
<th>Aggregated Reliability¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Frequency of crypts</td>
<td>2/576⁰</td>
<td>.91</td>
<td>.95</td>
</tr>
<tr>
<td>2. Frequency of nevi</td>
<td>2/310⁰</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td>3. Iris color</td>
<td>2/544⁰</td>
<td>.91</td>
<td>.95</td>
</tr>
<tr>
<td>4. Distinction of white dot rings</td>
<td>2/589⁰</td>
<td>.83</td>
<td>.91</td>
</tr>
<tr>
<td>5. Distinction of contractional furrows</td>
<td>2/589⁰</td>
<td>.91</td>
<td>.95</td>
</tr>
</tbody>
</table>

Note: ¹ The Spearman-Brown formula.
² All photos in the sample except those used in the scales and for definitions.
³ 51% of the total sample.
The parameter estimates from the four models are reported in Table 4 and the parameter estimates and 95% confidence intervals for the best fitting model are given in Table 5. The relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

Model 4, where the additive genetic, and non-shared environmental factors were constrained to be the same for both sexes, and the shared-environmental factor or the non-additive genetic factor were fixed at 0 for both sexes provided the most parsimonious explanation of the data for all five iris characteristics but eye color and distinction of white dot rings. The shared environmental parameters for males and females for all iris characteristics, although the relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

Model 4, where the additive genetic, and non-shared environmental factors were constrained to be the same for both sexes, and the shared-environmental factor or the non-additive genetic factor were fixed at 0 for both sexes provided the most parsimonious explanation of the data for all five iris characteristics but eye color and distinction of white dot rings. The shared environmental parameters for males and females for all iris characteristics, although the relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

Significant differences in intra-class correlations between the young (4–24 years old) and the old cohort (25–72 years old) were evident for crypts in the iris stroma ($r_{MZYOUNG} = .44; r_{MZOLD} = .07$), distinction of white dot rings ($r_{MZYOUNG} = .72; r_{MZOLD} = .90$), and iris color ($r_{MZYOUNG} = .83; r_{MZOLD} = .93$). No other significant differences between the intra-class correlations for the young and the old cohort were found.

As expected, the intra-class correlations for the age-matched randomly paired unrelated dyads (both the same and opposite sex dyads) were close to zero and non-significant for all five iris characteristics. These dyads were not included in the model fitting procedure.

**Sex Differences and Age Effects**

The parameter estimates from the four models are reported in Table 4 and the parameter estimates and 95% confidence intervals for the best fitting model are given in Table 5. The relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

Model 4, where the additive genetic, and non-shared environmental factors were constrained to be the same for both sexes, and the shared-environmental factor or the non-additive genetic factor were fixed at 0 for both sexes provided the most parsimonious explanation of the data for all five iris characteristics but eye color and distinction of white dot rings. The shared environmental parameters for males and females for all iris characteristics, although the relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

Model 4, where the additive genetic, and non-shared environmental factors were constrained to be the same for both sexes, and the shared-environmental factor or the non-additive genetic factor were fixed at 0 for both sexes provided the most parsimonious explanation of the data for all five iris characteristics but eye color and distinction of white dot rings. The shared environmental parameters for males and females for all iris characteristics, although the relative importance of genetic effects was the same in males and females for all iris characteristics, although the male specific parameter was significant for frequency of nevi. Nevertheless, the most parsimonious model for frequency of nevi suggested that there were no qualitative genetic differences between males and females and that the parameter estimates were the same.

**Table 2**

Means and Standard Deviations for Five Scales Describing Iris Characteristics in MZ and DZ Twins

<table>
<thead>
<tr>
<th>Scale / Iris characteristics</th>
<th>MZ ($n = 100$ pairs)</th>
<th>DZ ($n = 99$ pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>1. Frequency of crypts$^1$</td>
<td>2.14</td>
<td>0.90</td>
</tr>
<tr>
<td>2. Frequency of nevi$^1$</td>
<td>2.03</td>
<td>1.04</td>
</tr>
<tr>
<td>3. Iris color$^{2,3}$</td>
<td>2.32</td>
<td>0.90</td>
</tr>
<tr>
<td>4. Distinction of white dot rings$^1$</td>
<td>1.40</td>
<td>0.67</td>
</tr>
<tr>
<td>5. Distinction of contractional furrows$^{1,3}$</td>
<td>2.25</td>
<td>1.18</td>
</tr>
</tbody>
</table>

Note: $^1$ Five scale-steps.

<table>
<thead>
<tr>
<th>Scale / Iris characteristics</th>
<th>MZ$_M$</th>
<th>DZ$_M$</th>
<th>MZ$_F$</th>
<th>DZ$_F$</th>
<th>DZ$_OS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pairs</td>
<td>$n = 46$</td>
<td>$n = 27$</td>
<td>$n = 54$</td>
<td>$n = 33$</td>
<td>$n = 39$</td>
</tr>
<tr>
<td>1. Frequency of crypts</td>
<td>.69**</td>
<td>.04 ns</td>
<td>.84**</td>
<td>.24 ns</td>
<td>.59**</td>
</tr>
<tr>
<td>2. Frequency of nevi</td>
<td>.60**</td>
<td>.29*</td>
<td>.57**</td>
<td>.51**</td>
<td>.05 ns</td>
</tr>
<tr>
<td>3. Iris color</td>
<td>.87**</td>
<td>.26 ns</td>
<td>.86**</td>
<td>.43**</td>
<td>.89**</td>
</tr>
<tr>
<td>4. Distinction of white dot rings</td>
<td>.75**</td>
<td>.07 ns</td>
<td>.75**</td>
<td>.22 ns</td>
<td>.26*</td>
</tr>
<tr>
<td>5. Distinction of contractional furrows</td>
<td>.74**</td>
<td>.32*</td>
<td>.83**</td>
<td>.28 ns</td>
<td>.25 ns</td>
</tr>
</tbody>
</table>

Note: ** $p < .01$; * $p < .05$. 

Significant mean differences between MZ and DZ twins ($p < .05$).

Significant variance differences between MZ and DZ twins ($p < .05$).
younger (73%) (see Table 5). The change in fit between model 1 (constraining the parameter estimates to be the same for the older and the younger cohorts) and the sum of the chi-squares for models 2 and 3 (separate parameter estimates for the younger and older cohort), was significant, change in $\chi^2 = 7.30, p < .05$.

### Discussion

The heritabilities of five characteristics of the human iris were substantial and ranged from .51 to .90. Heritability for distinction of white dot rings was greater for an older (25–72 years) than a younger (4–24 years) cohort. The set of genes that influenced the iris traits as well as their relative importance was the same in males and females for all traits.

Candidate genes responsible for variation in iris texture may be different alleles of Pax6, and its downstream target genes, COX-1, COX-2, NF1, VEGF, Ezrin, IR 185/PTC, FOXC1, TIGR/GCLI/ Myocilin, UBL5, RIE1, BARX1, CB1, TRP-1, TRP-2, OCA1, OCA2, EVCL1, EVCL2, EVCL3, TYRP1 the P gene, and QNR-71 (Aksan & Goding, 1998; Anderson et al., 2002; Boissy & Nordlund, 1997; Chang et al., 1999; Damm et al., 2001; Friedman et al., 2001; Friedman et al., 2000; Gould & Walter 2000; Huang & Wenhui, 2000; Jaworski et al., 1997; Kim et al., 1999; Kivelä et al., 2000; Kulak et al., 1998; Lowings et al., 1992; Lu et al., 1998; Orlow & Brilliant, 1999; Oetting & King, 1999; Procella et al., 2000; Ragge et al., 1993; Rebbeck et al., 2002; Shen et al., 1996; Simpson & Price, 2002; Wang et al., 2001; Zehavi et al., 1986) and other genes that expressed in the iris (Wistow et al., 2002; Wistow, 2002). Mutations in the Pax6 gene lead to aniridia, a malformation of the eye, chiefly characterized by iris hypoplasia, and disturbances in the development of the lens, cornea and retina (Glaser, 1992; Gronskov, 2001). This gene is also recognized as one of the most important control genes for the CNS and its expression pattern in the iris and throughout the CNS has been reported elsewhere (Chalepakis et al., 1993; Jaworski, et al., 1997; Mansouri et al., 1994; Simpson & Price, 2002; Van Heyningen & Williamson, 2002). Most noteworthy is Pax6 expression in the amygdala and involvement in the production of dopamin and noradrenaline neurons (Jaworski & Wistow, 1997; Stoykova & Gruss, 1994). Family members with a mutation in Pax6 show high rates of unusual behavior including disinhibition, impulsive behavior, impaired social

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimates and Goodness of Fit Statistics for the Five Iris Characteristics of Interest</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Crypts</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>Nevi</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>Color</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>White dot rings</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>Cont. furrows</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Note: Best fitting model when comparing the models is in bold.

1 Modeled as $d^2$.
2 Modeled as $d'm^2$.

Model 1: all parameters freely estimated.
Model 2: $a^2$, $c^2$ or $d^2$, $e^2$ fixed at 0 all other parameters freely estimated.
Model 3: the same parameter values ($a^2$, $c^2$ or $d^2$, $e^2$) were required for male and females, $a'm^2$, $c'm^2$, $d'm^2$, $e'm^2$ fixed at 0.
Model 4: $c^2$ or $d^2$ fixed at 0 for both males and females, the same parameter values ($a^2$, $e^2$) were required for male and females, $a'm^2$, $d'm^2$ fixed at 0.
understanding and impaired verbal inhibition (Heyman et al., 1999). Thus, different alleles of Pax6 may be responsible for a potential genetic correlation between iris patterns and personality.

The heritability estimates for eye color in the present study and the Louisville twin study (Bito, 1997) differed somewhat from each other. The heritability for iris color in the Louisville study was 98% for both males and females. This high heritability is in line with recent linkage findings that suggest that much of the genetic variance for eye color is explained by the OCA2 genes and the P gene (Duffy et al., 2003; Rebbeck et al., 2002). In our study the heritability was 85% for both males and females when the shared environmental parameter was fixed to 0. However, the shared environmental parameter is significant, primarily due to relatively high female DZ and opposite sexed DZ correlations. One potential explanation for this unlikely finding could be assortative mating for eye color, which would lead to elevated DZ correlations. However, Hasstedt (1995) found little evidence for assortative mating for eye color nor were spouses similar for eye color in another sample (N. G. Martin, personal communication, March 21, 2003). Furthermore, an artifact in the model can occur in small samples for traits caused by a major single locus with a large non-additive genetic effect, which is plausible for eye color, this is apparently not completely negligible. The risk that the intra-class correlations do not reflect the true nature of the genetic effects in large samples is usually very small, but as can be seen in our sample for eye color, this is apparently not completely negligible. The results from the larger Louisville twin study in conjunction with the linkage findings lead us to believe that shared environmental effect is due to a sampling effect.

Another difference between the two studies was that the Louisville twin study found age differences in heritability and non-additive genetic effects rather than additive genetic effects could be observed was distinction of white dot rings (see Table 5). The patterns for the DZ correlation for the young and old cohort followed the pattern that was observed for eye color in the Louisville Twin Study, with smaller DZ correlations in the older cohort (rMZOUNG = .83 versus rMZOLD = .93), but the parameter estimates did not differ significantly. This probably reflects our relatively small sample size as well as the broad age range. Furthermore, the Louisville study was longitudinal, and could evaluate true age effects rather than cohort differences.

The only iris characteristic where significant age differences in heritability and non-additive genetic effects rather than additive genetic effects could be observed was distinction of white dot rings but not for iris color may be that the scale for distinctiveness of white dot rings records much smaller changes in quantities of pigmentation rather than the iris color scale. Even small changes in pigmentation that cover white dots may be undetectable, since much more pigment is needed in order to influence the rating for this iris characteristic. Thus, even very small changes in pigmentation that occur over time may decrease the similarity for white dots rings within the DZ twin pairs, and result in a greater heritability for white dot rings in the older cohort.

Our capability to observe sex and age differences in heritability was limited by the relatively small sample size. Nevertheless, it is striking that so few age and sex differences in parameter estimates for the iris characteristics were found. It may be that most iris characteristics generally reach their genetically influenced appearance before adulthood, and then do not change much over time. Further investigations assessing iris characteristics other than eye color in large longitudinal twin samples with more power than the present study need to be performed in order to confirm this assumption.

### Table 5

<table>
<thead>
<tr>
<th>Scale/Iris characteristics</th>
<th>Additive genetic variance a² (95% CI)</th>
<th>Non-additive genetic variance d² (95% CI)</th>
<th>None-shared environmental variance e² (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Frequency of crypts¹</td>
<td>66% (55–75%)</td>
<td>—</td>
<td>33% (25–50%)</td>
</tr>
<tr>
<td>2. Frequency of nevi¹</td>
<td>58% (45–68%)</td>
<td>—</td>
<td>42% (32–56%)</td>
</tr>
<tr>
<td>3. Iris color¹</td>
<td>51% (29–81%)</td>
<td>34% (3–55)²</td>
<td>15% (11–21%)</td>
</tr>
<tr>
<td>4. Distinction of white dot rings¹</td>
<td>0% (0–75%)</td>
<td>78% (15–84%)</td>
<td>22% (16–31%)</td>
</tr>
<tr>
<td>Young cohort (4–24 years old)</td>
<td>73% (61–82%)</td>
<td>—</td>
<td>27% (19–39%)</td>
</tr>
<tr>
<td>Old cohort (25–72 years old)</td>
<td>90% (80–94%)</td>
<td>—</td>
<td>10% (5–20%)</td>
</tr>
<tr>
<td>5. Distinction of contractional furrows¹</td>
<td>78% (70–84%)</td>
<td>—</td>
<td>22% (16–30%)</td>
</tr>
</tbody>
</table>

Note: ¹ No quantitative or qualitative sex difference in the estimates
² d² was modeled as c²
The present study analyzed the same material as Burkhardt (1992). However, the results presented in the earlier German study did not allow for estimation of the relative importance of genetic effects. Furthermore, the rater was not blind to zygosity. Therefore, we chose to repeat the rating process, with raters blind to zygosity and using standardized scales that covered the whole range of variability for each iris characteristic in the sample. Another advantage was that two independent judges rated each photograph, which made it possible to test inter-rater reliability. Inter-rater reliability was substantial, greater than .90 for each of the five scales. The most important factor contributing to the high inter-rater reliability was probably the use of the rating catalog and references photos which exemplified in great detail what the raters should do when they were confronted with photos that fell between two scale steps. The second most important factor was probably feedback given to the raters concerning their inter-rater reliability score after each rating session. Furthermore, all pictures that not all raters judged the same way were discussed with the test leader. This led to a higher score in the next rating session. Despite these methodological differences, there were striking similarities in the results of Burkhardt (1992) and the present study. The iris characteristics that were estimated to have highest heritability in the present study were also considered to be highly heritable by Burkhardt (i.e., eye color, crypts in the main stroma leaf, distinction of white dots and distinction of contractional furrows).

The random influences on how the iris tissue grows and develops over time were combined with measurement error in the non-shared environmental parameter (e²), which ranged from .10–.42. The frequency of nevi and crypts in the iris were most influenced by random factors (Table 5). For most of the characteristics, we believe that measurement error is the most important source of e². In the case of nevi, however — which has the largest e² — developmental changes which can reflect disease later in life (Habour et al., 1995), may be another source of e².

**Methodological Implications for Behavioral Research**

The main purpose of the present study was to estimate the heritability of the iris features that had shown the strongest relationship with personality (Bruno, 1990; Larsson, 1998). An additional purpose was to develop a method for training raters to secure the reliability of the iris tissue texture estimate. Given the high inter-rater reliabilities and heritabilities for the iris characteristics reported here, we suggest that it may be fruitful to perform additional studies that replicate the associations found by Larsson (1998) and Bruno (1990), but that include both personality and iris data collected from twins in order to evaluate whether the associations reflect genetic correlations among the traits. Due to the ethnic stratification of eye color and personality (and putatively also of other iris characteristics), will it be necessary to conduct within-family tests of a potential association.

If the genes that are expressed in the iris tissue explain substantial amounts of the genetic variation in iris characteristics, and if there is a substantial genetic correlation between personality and iris characteristics, then the iris characteristics may be considered as biomarkers of personality. Such findings may give the old expression “it was all in the eyes” a new meaning in the future.

**Acknowledgment**

This study was supported by funds from the Department of Social Sciences at Örebro University in Sweden and The Swedish Foundation for International Cooperation in Research and Higher Education. We are grateful to Dr Angelica Burkhardt for preparing the data material, Anna Törnblom and Ulrika Sandberg for their accurate work as raters, and all the twins as well as the unrelated participants for their cooperation.

**References**


