Individual differences in visual perception and memory

Colizoli, O.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Abstract
Grapheme-color synesthesia is defined as the experience of color in relation to letters, words and numbers. Grapheme-color and linguistic-color synesthesia have a genetic component. Still, language (the most common synesthetic inducer) is acquired through an interaction between genes and the environment. Similarly, grapheme-color synesthesia is likely shaped by both genetic and environmental factors. We compared non-synesthetic relatives of grapheme-color synesthetes to a group of matched controls in a reading-in-color training paradigm. All participants read specially prepared books in which four high-frequency letters were paired with four high-frequency colors. Magnetic resonance images were acquired before and after training, including diffusion tensor imaging scans. Behavioral results replicated our previous findings of learned automatic letter-color associations by reading in color tested with a Stroop task (Colizoli et al., 2012). Imaging results showed significant brain activation related to congruency in regions known to be involved in (authentic) grapheme-color synesthesia, including inferior occipital lobe near V4, inferior and superior parietal lobe, precentral gyrus, and insular cortex. There was no evidence of increased acquisition of the letter-color associations in the behavior of the relatives of synesthetes, however, neural activation was found to be significantly different between groups in regions associated with synesthesia, including near area V4. The post-training Stroop effect was (marginally) correlated with activation in color regions (near area V4) in response to seeing the trained letters (presented in black) compared to untrained letters. Using region of interest analyses, differences between relatives and controls in brain structure were found in grey and white matter. Controls had larger grey matter volume in the right angular gyrus, whereas relatives had increased structural connectivity measured with fractional anisotropy in the left superior frontal lobe. We conclude that activation related to training synesthetic associations as well as pre-existing brain structure are significantly different in the relatives of synesthetes compared to controls and most notably in multi-sensory brain regions. These differences in brain networks in conjunction with the lack of any measurable differences in behavior between groups suggests differential network processing with identical behavioral output.

Introduction
The development of synesthesia is determined by an interaction between genes and environment, because the presence of synesthesia runs in families (Tomson et al., 2011; Asher et al., 2009) and the majority of synesthetic inducers (e.g. language) are culturally transmitted (Barnett et al, 2008; Rich et al. 2005). Synesthetes show differences in both brain function and structure (for a review see Rouw et al., 2011) that provide a window into understanding how and where ‘synesthetic’ genes may be expressed in the brain. Novich et al. (2011) showed that 22 different forms of synesthesia can be classified into five subgroups, suggesting independent genetic and neural bases: colored sequences, colored music, non-visual sequela, spatial sequences, and colored sensations. To date, two studied have investigated the candidate regions on the genome related to colored hearing synesthesia (Asher et al., 2009) and colored
sequence synesthesia (Tomson et al., 2011). Both studies suggested multiple modes of inheritance as well as locus heterogeneity. This is not surprising given the idiosyncratic nature of synesthesia. There is evidence that genes provide a predisposition for developing the trait of synesthesia, while the type of synesthesia is not inherited but determined by the environment (Rouw et al., 2011). Relatives share a percentage of genetic code as a function of their relationship (Falconer, 1965). Therefore, researching the neural basis of non-synesthetic relatives of synesthetes can shed light on the probabilities and underlying mechanisms associated with developing (the trait and type of) synesthesia, or subgroup of synesthesia. There are many questions to answer concerning the role of the environment in shaping synesthetic associations and whether specific genes are necessary or sufficient for the development of synesthesia.

Currently, it is unknown to what extent synesthetic associations are ‘picked up’ directly from the environment. Grapheme-color synesthesia (the experience of color in relation to letters, words and numbers) is relatively common and well-characterized in the literature. In at least one sample of grapheme-color synesthetes, specific letter and number-color mappings came directly from the same set of refrigerator magnets that each individual played with as a child (Witthoft & Winawer, 2013). A large-scale analysis did not, however, reveal connections between children’s books in Australia and the letter-color mappings of adult synesthetes in that country (Rich et al., 2005). The probability of finding two synesthetes of the same type with the exact same set of inducer-concurrent mappings is extremely low. Each synesthete has a unique set of letter-color mappings that remain highly consistent over time.

The specific mappings are not always directly related to the environment, but research has shown that correlations exist between the environment and synesthetic mappings at a meta-level; for instance, higher frequency graphemes tend to induce higher frequency colors terms (Simner et al., 2005) and have more luminance and saturation (Smilek et al., 2007). In lexical-gustatory synesthesia, all food words taste like the foods they describe, and a complex relationship between these food words, semantics and phonology determines the entire set (food and non-food words) of word-taste associations for each individual synesthete (see Chapter 3). Taken together, the evidence suggests that the specific synesthetic mappings are learned early in life and largely dependent on the environment (i.e. language and culture).

Whether synesthetic experiences are the result or the cause of structural brain differences remains to be shown. Most likely individuals develop synesthesia because of a genetic predisposition for forming semantic and perceptual connections (perhaps via expression of genes related to neuronal migration; Asher et al., 2009), and the environment in which they learn the concepts and categories of the world. These specific associations (i.e. the content of the experiences) become ‘hard-wired’ in brain structure over time. It is well known that the brain is more plastic in childhood, but recent research shows that experience-dependent brain plasticity is evident even in adulthood (Draganski et al., 2004; Colcombe et al., 2006; Boyke et al., 2008; Driemeyer et al., 2008; Draganski et al., 2006; Scholz et al., 2009; Tang et al., 2010; Kwok et al., 2011). Is it possible to form synesthesia in adulthood by training? Would relatives
of synesthetes be more likely to experience synesthesia than controls after training? (For an in-depth discussion about why we should bother trying to train synesthesia at all, see Chapter 7)

In order to address these questions, we investigated the interaction between synesthetic genes and the environment by comparing 11 (non-synesthetic) relatives of grapheme-color synesthetes and 11 matched controls in a synesthetic training paradigm. In addition to pre-existing differences in brain structure and function, we tested the effect of reading in color using functional magnetic resonance imaging (fMRI), voxel-based morphometry (VBM), and diffusion tensor imaging (DTI).

Before reading began, each participant chose which book(s) he or she would like to read and was assigned a unique set of four letters and four colors (Figure 5.1) based on pre-existing letter-color preferences. Before reading, T1-weighted structural images were collected in the MRI scanner and the Stroop and crowding tasks were administered in the computer lab. After reading, we again collected T1-weighted structural images and participants completed the Stroop and crowding tasks while being scanned as well as in the computer lab. Importantly, we included a baseline condition in the Stroop task consisting of letters that had not been in color in the books as a within-subjects control measure in order to compare trained versus untrained letters. In addition a color localizer was used to test for brain activation related to both veridical and ‘synesthetic’ color.

To date, this is the first study of its kind in several respects: The brains of relatives of synesthetes were compared to a group of matched controls. Related to training ‘synesthetic’ associations by reading in color, effects on brain function and structure were investigated. A complex relationship could be characterized related to the interaction between relatives and controls in terms of both behavior and neural mechanisms involved in training these associations.

Materials and Methods
Participants
Relatives of synesthetes, who did not have any form of synesthesia themselves, were recruited by contacting synesthetes from our participant database and asking them if they had any relatives who would be interested in participating in a study about reading in color. Eleven adult relatives of synesthetes (8 female, \( M = 24.73 \text{ years}, SD = 2.32 \)) and 11 matched controls (8 female, \( M = 25.18 \text{ years}, SD = 2.64 \)) took part in the study (entire sample: \( M = 24.95 \text{ years}, SD = 2.43 \), range = 22 – 30 years). Three pairs of brothers and sisters were in the relative group. The familial relationships of the participants to their synesthetic relatives are given in Table 5.1. Controls were matched for age, sex, education, and handedness. Participants were screened for synesthesia, dyslexia, and attention-deficit disorder (ADD), in addition to the standard MRI screening protocol. Participants were tested for color blindness using the Ishihara test for color blindness (Ishihara, 1936). The native language of each participant was Dutch.

The grapheme-color synesthesia of the synesthetic relatives of our participants was verified with a standardized synesthesia battery (Eagleman et al., 2007) or with a questionnaire test-
retest paradigm (Baron-Cohen et al., 1987; Baron-Cohen et al., 1993). Synesthetes were compensated financially for their participation.

**Table 5.1 Familial relationships of the participants to their synesthetic relatives.** Presented are the participants’ gender, the participants’ relationship to their synesthetic relative(s), the estimated percentage of genes shared with their synesthetic relative(s) (% Syn. Genes), and the side of the family of the relationship. The three pairs of brothers and sisters who participated in the current study are marked with matching asterisks (*, **, and ***).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Relationship to Synesthete(s)</th>
<th>% Syn. Genes</th>
<th>Side of Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>3*</td>
<td>F</td>
<td>Half sister</td>
<td>25</td>
<td>Father</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>Sister, Daughter</td>
<td>50</td>
<td>Mother &amp; Father, Mother</td>
</tr>
<tr>
<td>6*</td>
<td>M</td>
<td>Half brother</td>
<td>25</td>
<td>Father</td>
</tr>
<tr>
<td>8***</td>
<td>M</td>
<td>1st Cousin</td>
<td>12.5</td>
<td>Father</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Sister</td>
<td>50</td>
<td>Mother &amp; Father</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>Sister, Daughter</td>
<td>50</td>
<td>Mother &amp; Father, Father</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>Sister</td>
<td>50</td>
<td>Mother &amp; Father</td>
</tr>
<tr>
<td>14**</td>
<td>F</td>
<td>Sister</td>
<td>50</td>
<td>Mother &amp; Father</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>Daughter</td>
<td>50</td>
<td>Mother</td>
</tr>
<tr>
<td>18**</td>
<td>M</td>
<td>Brother</td>
<td>50</td>
<td>Mother &amp; Father</td>
</tr>
<tr>
<td>22***</td>
<td>F</td>
<td>1st Cousin</td>
<td>12.5</td>
<td>Father</td>
</tr>
</tbody>
</table>

This experiment was approved by the ethical committee of the Department of Psychology at the University of Amsterdam. All participants were informed that they could terminate their participation at any time and gave written informed consent before participating in the research. Participants were rewarded financially for participating in this study.

**General Procedure**

Before testing began, participants completed a questionnaire asking them to rate their preferences for the 16 pairs of the four possible letter-color combinations and chose one or more books to read from a list of books. Participants completed both a behavioral and MRI testing session before and after reading the colored books. The behavioral session consisted of a ‘synesthetic’ Stroop task, a perceptual crowding task, a measure of reading using an eye tracker (data not reported here), and a set of questionnaires. During the MRI sessions, functional brain activation and brain structure were measured in order to quantify changes induced by reading in color in activation, grey matter volume and white matter coherence within each voxel.

**Books**

All participants read specially prepared books that contained four high-frequency letters in four high-frequency colors. The colored letters were ‘a’, ‘e’, ‘n’, and ‘r’ (only lower-case). The colors were red, orange, green and blue. The specific pairs of letter-color combinations varied between participants and were based on a questionnaire where they reported their preferences for each letter-color pair (see Preference for letter-color pairs).
The permission to reprint and use the books, along with the digital copies of the original books, were given to us by the Publisher Nijgh & Van Ditmar (www.nijghenvanditmar.nl). All books were in the Dutch language. The font of each book was converted to Arial Black (10 pt). The letters were changed into color using a customized Visual Basic macro application (Colizoli et al., 2013). The content of the books was not altered in any way. The word and character counts for each book were noted.

The instructions for reading included a statement asking participants to be honest about how much they read, even if they did not manage to finish a book. If this was the case, participants marked the spot in the book where they stopped reading before the testing session. The word and character counts up to this point were included in the total word and character count for the participant. During the post-reading testing session, participants were asked to give an account of the plot of the book(s) that they read and also their opinion of the book(s) in general. Everyone responded with a correct description of the events in the book. There was no reason to believe that any of the participants had not read what they claimed to have read.

Preference for letter-color pairs
The preference for letter-color pairs and the frequency of the letters may interact. In order to test for this and make sure it was balanced across participants, we used a procedure that was identical to the one reported in Colizoli et al., (2012). Before testing began, participants were given a questionnaire which asked them to indicate their preference for the four letters (‘a’, ‘e’, ‘n’, and ‘r’) in all four of the colors used (red, orange, green, and blue; 16 questions, 5-pt. Likert scale). Preferred letter-color pairs were defined as a score of ‘3’ or higher, while non-preferred letter-color pairs were defined as a score of ‘3’ or lower (a score of ‘3’, i.e. a neutral preference, could be included in either category). In order to separate any effects of color preference from letter frequency, relatives of synesthetes were randomly assigned to two groups: The first group was given two of their preferred letter-color pairs for the highest frequency letters (‘e’ and ‘n’) and two of their non-preferred letter-color pairs for the lowest frequency letters (‘a’ and ‘r’). The second group was given two of their non-preferred letter-color pairs for the highest frequency letters and two of their preferred letter-color pairs for the lowest frequency letters. Each matched control was assigned to the same preference group as their matched counterpart (from the group of relatives) but received unique letter-color pairs based on their individual preferences.

Behavioral testing procedure
Before and after reading, participants were given a Stroop task, a crowding task and were asked to read text on the screen. After these behavioral tasks were completed, participants filled out several questionnaires. The order of the behavioral tasks was counterbalanced between participants but remained the same for each individual across testing sessions. Each control participant received the same procedure as the participant he or she was matched to in the relatives of synesthetes group. Each behavioral session took approximately one hour to complete. Participants were seated 50 cm in front of computer monitor. This distance remained constant by the use of a chin rest. All stimuli were presented on a PC with Presentation (version 14; www.neurobs.com) on a 20-inch VGA monitor. The screen resolution was 1280 x 1024 pixels. All responses were recorded with a USB keyboard.
**Stroop task**
This task was modeled after Meier and Rothen (2009). In the ‘synesthetic’ Stroop task, participants were shown one of eight letters (‘a’, ‘b’, ‘e’, ‘g’, ‘k’, ‘n’, ‘r’, or ‘t’) on screen in black, then changed rapidly into one of four colors (red, orange, green or blue). Participants were instructed to indicate this color as fast and accurately as possible. They answered via key press on one of four buttons. There were three letter conditions: 1) the four trained letters (‘a’, ‘e’, ‘n’, and ‘r’) presented in colors congruent with the colors within the books, 2) the four trained letters (‘a’, ‘e’, ‘n’, and ‘r’) presented in colors incongruent with the colors within the books, and 3) the untrained condition consisted of four untrained letters (‘g’, ‘t’, ‘k’, and ‘b’) that were always presented in black within the books. The Stroop effect is the difference between incongruent and congruent trials. Reaction times and accuracy were recorded as dependent variables. The Stroop task was presented in three blocks, with a short self-paced break in between blocks. The congruent, incongruent and untrained trials were randomized within each block, with a total of 288 trials (96 congruent, 96 incongruent, and 96 untrained trials).

A single trial consisted of one of four letter presented in black for 200 ms, followed by the same letter presented in one of the four colors until a response was made (reaction time) or 2000 ms (missed trial). The inter-trial interval (a fixation cross) was jittered (a random integer between 1000 and 2000 ms). All stimuli were presented in the center of the screen on a white background. The average visual angle of letters was 2.00° (56.6 pixels).

Before the Stroop task began, participants learned which buttons corresponded to the four colors used (red, orange, green and blue). Colored squares were presented on screen until a response was made. Participants were instructed to be as fast and accurate as possible and received feedback on each trial. Participants completed one round of 192 trials and were monitored during the task to see if they were looking at their fingers before responding. If this was the case, they were asked to do another set of button-color training. Colored stickers were placed on the keyboard for both the button training and Stroop task.

The task was slightly modified when presented in the MRI scanner: Two runs of the Stroop task were presented in the post-reading MRI testing session. Each run lasted approximately seven minutes. In each run, 72 trials were presented in random order (24 trials per letter condition). The colored letters remained on screen for a fixed duration of two seconds. This task was an event-related design and the inter-trial interval was jittered in the following way: 50% of ITIs = 2 s, 33.3% of ITIs = 4 s, 5.6% of ITIs = 6 s, 5.6% of ITIs = 7 s, 5.5% of ITIs = 8 s. All other features of the task remained identical.

**Crowding task**
This task was modeled after Hubbard et al. (2005). The crowding stimuli formed a cross, consisting of a target letter (‘a’, ‘e’, ‘n’ or ‘r’) surrounded by one of the remaining three letters (along the cardinal axes). Participants were instructed to identify the middle letter in the group of letters appearing to either the left or the right of fixation. Participants were instructed to fixate while the letters appeared and their eye-movements were recorded with an eye-tracker (EyeLink; www.sr-research.com). We included a untrained letter condition (‘m’, ‘s’, ‘u’, ‘w’), in
which visual pop-out would not be expected to occur since none of the four letters in that condition were colored in the book.

The four flanking letters were always the same as each other, but never the same as the target letter (this information was withheld from participants). There were 12 combinations of letters in each letter condition. Each of these 12 combinations of letters was presented on the left and right side of fixation, making 24 total trials per letter condition. These 24 trials were repeated three times per letter condition in random order. The trained and untrained conditions were presented in alternating order in separate blocks (three blocks each). There were 72 trials in each letter condition (144 total trials). The participants took a short self-paced break in between each block. The order of the trained and untrained conditions was counterbalanced between participants, but remained the same during the pre- and post-reading testing sessions for each participant. Participants completed 24 practice trials on a different set of letters (‘c’, ‘k’, ‘o’, ‘u’) before beginning the task. Accuracy was recorded as the dependent variable.

A single trial began with a fixation cross for a jittered duration (random integer between 500 and 3500 ms) in order to prevent anticipatory eye movements. After this fixation period, the crowding stimuli were presented on either the left or right side of fixation for 200 ms. After the crowding stimuli, a blank screen was presented for 250 ms before a four-alternative-forced-choice (4-AFC) response screen appeared. This 4-AFC screen went away upon response or 2000 ms (missed trial). All stimuli were black on a white background. The vertical size of the letters in the crowding task subtended a visual angle of 1.02 on average (38.5 pixels), and the average center-to-center spacing of the letters subtended a visual angle of 1.53 on average (57.7 pixels).

The task was slightly modified when presented in the MRI scanner: Two runs of the crowding task were presented in the post-reading MRI testing session. Each run lasted approximately seven minutes. In each run, six blocks of each letter condition were presented. Each block consisted of four trials. A total of 48 trials were presented in each run (24 trials per letter condition). The task was presented as a blocked-design with 16-second stimuli blocks and 16-second periods of rest between blocks. The inter-trial interval was set at a fixed duration of 1550 ms. The 4-AFC screen remained on screen for a fixed duration of two seconds. All other features of the task remained identical. Eye movements were also recorded in the MRI scanner.

**Questionnaires**

Several questionnaires were administered during each behavioral testing session. In the pre-reading testing session, participants completed the a general screening form, the first part of a test of synesthetic consistency, the Vividness of Visual Mental Imagery Questionnaire (VVIQ; Marks, 1973; Cui et al., 2007), and an objective test of visual mental imagery ability. In the post-reading testing session, participants completed the second part of the synesthetic consistency test, the Projector-Associator Questionnaire (Rouw & Scholte, 2007), and a general questionnaire about their reading experience (Colizoli et al., 2012).

In order to objectively rule out grapheme-color synesthesia within the sample, we administered a paper version of the synesthetic test of consistency, a test-retest paradigm. The list consisted of all letters (A-Z), numbers (0-9), and days of the week. The post-reading version of this list was
exactly the same as the pre-reading version except the order of the items had been shuffled in order to prevent memory strategy or context effects. For the test of consistency, we asked participants to openly indicate which color each item would have if they would have a color (and they had to choose a color).

Brain imaging procedures
All participants were screened for MRI-related health risks using a standardized questionnaire. Each participant completed a MRI session before and after reading the colored books. In each of these sessions, two T1-weighted structural images and four runs of DTI were acquired. The functional tasks differed between testing sessions: a visual word form area localizer and a retinotopic mapper were administered during the pre-reading MRI session (these data are not reported here). In the post-reading MRI session, the Stroop and crowding tasks were administered in addition to a (synesthetic) color localizer. Functional and structural data were collected using a fixed order of runs across participants (this order was optimized during piloting in order to minimize fatigue while in the scanner). All functional runs began and ended with an 18-second baseline period. Each MRI session took 1.5 hours of scanning time.

fMRI acquisition and data analysis
Scans were acquired on a Philips 3 Tesla Achieva TX scanner, located at the Spinoza Center, Amsterdam, the Netherlands. Whole brain gradient-echo echo-planar imaging (EPI) measurements (voxel size = 3 x 3 x 3 mm, repetition time [TR] = 2000 ms, echo time [TE] = 27.63 ms, flip angle [FA] = 76.1°, FOV = 240 x 240, matrix = 80 x 80, slice thickness = 3 mm, slice gap = 0.3 mm, 38 slices per volume, sensitivity encoding factor of 2) were acquired to measure blood oxygen level-dependent (BOLD) magnetic resonance images with a 32-channel SENSE head coil. Each functional run of the Stroop task consisted of 225 volumes and lasted 7.5 minutes. Each functional run of the color localizer consisted of 330 volumes and lasted 11 minutes. Each functional run of the crowding task consisted of 190 volumes and lasted 6.3 minutes.

Analyses of the MRI images were carried out using FMRIB Software Library (FSL) version 5.0.4, Oxford, UK: http://www.fmrib.ox.ac.uk/fsl (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012). Statistical analyses were conducted using FSL’s fMRI Expert Analysis Tool, (FEAT version 6.00). Preprocessing steps included pre-whitening (FILM algorithm), spatial smoothing (a 5mm Gaussian kernel of full-width at half-maximum), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma = 50$ s). Voxels belonging to brain tissue were extracted from non-brain tissue voxels using the Brain Extraction Tool (BET; Smith, 2002). No runs were discarded due to motion (< 1 mm in all directions) or other artifacts.

In the first-level analysis, the time course of each run was convolved with the double gamma hemodynamic response function and tested with an uncorrected voxel threshold of $p < .05$. For the Stroop task (event-related) the temporal derivatives of each EV were included as confound regressors. Resulting contrast images were linearly registered to the anatomical structure using FLIRT with 7 degrees of freedom and the full search space (Jenkinson and Smith, 2001;
Jenkinson et al., 2002; Greve & Fischl, 2009), then spatially normalized to the T1-weighted MNI-152 stereotaxic space template (2 mm) using FNIRT with 12 degrees of freedom and the full search space (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FNIRT).

The higher-level analysis was carried out using FLAME (FMRIB's local analysis of mixed effects) stages 1 and 2 with automatic outlier detection (Beckmann, Jenkinson & Smith, 2003; Woolrich et al., 2004; Woolrich, 2008). Z-statistic (Gaussianised T/F) images were thresholded using clusters determined by $Z > 2.3$ and a corrected cluster significance threshold of $p = .05$, controlling the family-wise error rate (Worsley, 2001).

**VBM acquisition and data analysis**

In the first and last run of each MRI session, a T1 anatomical scan was acquired (four T1 volumes per participant, voxel size = $1 \times 1 \times 1$ mm, TR = 8229 ms, TE = 3.77 ms, FA = 8°, FOV = 256 x 256, matrix = 256 x 256, slice thickness = 1 mm, no slice gap, 160 slices per volume, 1 volume was acquired that lasted approximately 5 minutes) so that functional images could be registered to native anatomical space and normalized to the Montreal Neurological Institute (MNI) standard space.

T1-weighted structural data was analyzed using an optimized voxel-based morphometry protocol (Good et al., 2001; Douaud et al., 2007) as part of FSL. First, structural images were brain-extracted and grey matter-segmented before being registered to the T1-weighted MNI-152 standard space using non-linear registration (Andersson et al., 2007a; 2007b). The resulting images were averaged and flipped along the x-axis to create a left-right symmetric, study-specific grey matter template. Second, all native grey matter images were non-linearly registered to this study-specific template and ‘modulated’ to correct for local expansion (or contraction) due to the non-linear component of the spatial transformation. This study-specific template was made separately for the analyses of comparing sessions and comparing groups. When comparing groups, data was collapsed across sessions. The modulated grey matter images were then smoothed with an isotropic Gaussian kernel with a sigma of 4 mm. Finally, voxel-wise t-tests and correlation analyses (GLM) were implemented using the Threshold-Free Cluster Enhancement (Smith & Nichols, 2009) option in Randomise (Nichols & Holmes, 2002), a permutation-based non-parametric test (Anderson & Robinson, 2001) with 25,000 permutations, correcting for multiple comparisons across space (family-wise error rate = 5%).

**DTI acquisition and data analysis**

Fractional anisotropy was calculated on the basis of the acquisition of diffusion-weighted spin-echo EPI measurements (voxel size = $2 \times 2 \times 2$ mm, repetition time [TR] = 6310 ms, echo time [TE] = 73.36 ms, flip angle = 90°, FOV = 224 x 224, matrix = 112 x 112, slice thickness = 2 mm, no slice gap, 60 slices per volume, sensitivity encoding factor of 2). Diffusion was measured in 32 noncollinear directions. The start of each run was preceded by the acquisition of a non-diffusion-weighted volume for purposes of registration for motion correction. Four runs of diffusion-weighted images were acquired in each MRI session (8 in total). Each DTI run consisted of 34 volumes and lasted approximately 4 minutes.
A voxel-wise statistical analysis of FA data was carried out using Tract-Based Spatial Statistics (TBSS; Smith, 2006), part of FSL. First, FA images were created by fitting a tensor model to the raw diffusion data using FDT, and then brain-extracted using BET (Smith, 2002). All participants’ FA data were then aligned into a common space using the nonlinear registration tool FNIRT (Andersson et al., 2007a; 2007b), which uses a b-spline representation of the registration warp field (Rueckert, 1999). Next, the mean FA image was created and thinned to create a mean FA skeleton that represents the centers of all tracts common to the group. This study-specific skeleton was made separately for the analyses of comparing sessions and comparing groups. When comparing groups, data was collapsed across sessions. Each participant’s aligned FA data was then projected onto this skeleton. Finally, voxel-wise t-tests and correlation analyses (GLM) were implemented using the Threshold-Free Cluster Enhancement (Smith & Nichols, 2009) option in Randomise (Nichols & Holmes, 2002), a permutation-based non-parametric test (Anderson & Robinson, 2001) with 25,000 permutations, correcting for multiple comparisons across space (family-wise error rate = 5%).

**Regions of interest (ROIs)**
Whole-brain analyses of grey and white matter were conducted in order to test for differences between groups and differences between testing sessions. No significant results were found in the whole-brain analyses of either grey or white matter structure. Therefore, ROI analyses were performed. Different ROIs were used for grey and white matter analyses because of inherent differences in the nature of the tissue types. More specifically, the BOLD-signal originates from grey matter, and not white matter. Therefore, functionally defined ROIs may be better suited for analyses of grey matter structure analyses than for analyses of white matter structure (although there is white matter present in these functionally defined areas, generally it is not part of a major fiber tract). In addition, regions of interest in major white matter tracts were based on previous results that found significant differences between synesthetes and controls in structural connectivity (Rouw & Scholte, 2007). Similarly, white matter tracts do not contain much grey matter.

To test for group differences in grey matter, a mask was created *a posteriori* from the regions where a significant interaction in fMRI activation during the Stroop task was obtained between groups and conditions (trained vs. untrained and congruent vs. incongruent). Significant clusters of activation were thresholded at a cluster-based level of $Z > 2.3$.

Four regions of interest in the analysis of fractional anisotropy of white-matter tracts were defined *a priori* based on clusters of increased FA found in developmental synesthetes (Rouw & Scholte, 2007). These regions were: left and right superior frontal cortex underneath the central sulcus, left superior parietal cortex, and left inferior temporal cortex. Based on the coordinates given in Rouw and Scholte (2007), a spherical ROI was created with a diameter of 15 mm. T-tests were conducted on the averaged FA in each ROI and the number of tests was corrected using the Bonferroni method.

**Color localization**
A color localizer was used to test for voxels that respond to the presence of real and synesthetic color (modeled after: van Leeuwen, 2010). The localizer was a 16-second blocked design with
three conditions: 1) trained letters presented in black (‘a’, ‘e’, ‘n’, and ‘r’), 2) untrained letters presented in black (‘o’, ‘z’, ‘u’, and ‘w’), and 3) untrained letters presented in eight distinct colors (‘c’, ‘m’, ‘v’, and ‘s’; each of these letters was presented in all eight colors). Conditions were presented in pseudo-randomized blocks and each condition was presented six times. Within each block, stimuli were randomized, and each stimulus was presented for 500 ms with an inter-trial interval of 500 ms. During the 16-second rest period, a fixation cross remained on screen. One run of the localizer was presented and lasted approximately 10 minutes. Participants passively viewed the stimuli. The effect of color was defined by the contrast: colored letters > untrained letters in black (condition 3 > 2). ‘Synesthetic’ color was defined by trained letters in black versus untrained letters in black (condition 1 vs. 2, negative and positive differences).

Results
The significant level for all statistical analyses was set at $\alpha = 0.05$ unless otherwise stated. All t-tests were two-tailed unless otherwise stated. All reaction time (RT) data was based on correct trials only and is reported in milliseconds. Effect sizes reported for ANOVAs are partial eta squared ($\eta^2$). Non-parametric (Spearman) correlations are reported as $r_s$. Coordinates for brain regions are in Montreal Neurological Institute (MNI) space.

Behavioral results

Reading
Participants ($N = 22$) read an average of 80,398.50 words ($SD = 38,278.92$) and 381,008.00 characters ($SD = 182,383.02$) within an average of 20 days ($SD = 11.05$). There were no differences between relatives of synesthetes and controls in terms of the amount of words, characters or books read. An example of the colored text is illustrated in Figure 5.1.

Figure 5.1 An example of colored text used in the reading in color training paradigm. Four letters (a, e, n, r) were paired with four colors (red, orange, green, blue) based on the individual preferences for letter-color pairs.
Stroop task (outside of scanner)
A modified version of the Stroop task was administered in the computer lab before and after the specially prepared colored books were read. We were interested in the difference between the congruent and incongruent conditions. Data are given in Table 5.2. A repeated measures ANOVA was carried out for RT and accuracy on 2 factors: testing session (2 levels) and congruency (2 levels). RTs greater than 2.5 standard deviations and less than 150 ms were removed from the RT data per participant and condition (2.9% trials removed). In RT, there were significant main effects found for session, $F(1,21) = 7.49, p = 0.012, \eta^2 = 0.263$, and congruency, $F(1,21) = 12.97, p = 0.002, \eta^2 = 0.382$. The interaction between session and congruency was significant, $F(1,21) = 4.75, p = 0.041, \eta^2 = 0.184$ (Figure 5.2A). Post-hoc t-tests showed that participants responded to incongruent trials significantly slower than congruent trials after reading, $t(21) = 4.68, p < 0.001$, but not before reading, $t(21) = .98, p = 0.336$.

Table 5.2 Stroop task behavioral data. Mean reaction times and accuracy (and standard deviations) on the Stroop task before and after reading in color.

<table>
<thead>
<tr>
<th>RT (ms)</th>
<th>Pre Reading</th>
<th>Post Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>640.22 (81.26)</td>
<td>604.59 (72.11)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>647.08 (72.20)</td>
<td>630.43 (75.79)</td>
</tr>
<tr>
<td>Baseline</td>
<td>642.42 (84.30)</td>
<td>619.59 (72.30)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy (%)</th>
<th>Pre Reading</th>
<th>Post Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>95.82 (2.77)</td>
<td>94.55 (4.16)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>94.55 (3.69)</td>
<td>93.59 (3.70)</td>
</tr>
<tr>
<td>Baseline</td>
<td>95.41 (3.43)</td>
<td>94.41 (3.59)</td>
</tr>
</tbody>
</table>

In accuracy, a marginal effect was found for session, $F(1,21) = 3.73, p = 0.067, \eta^2 = 0.151$, and a significant effect was found for congruency, $F(1,21) = 4.80, p = 0.040, \eta^2 = 0.186$. No interaction between testing session and congruency was found ($F < 1$; Figure 5.2B), and this was probably due to the fact that accuracy in all conditions was very high. From now on, we will only refer to the Stroop effect in the RT data.

No differences between relatives of synesthetes and controls were found for the Stroop data in RT or accuracy.
A modified version of the behavioral Stroop task was administered in the MRI scanner after the specially prepared colored books were read. The behavioral log files of one participant was lost due to a technical error (control group), therefore, \( N = 21 \). Incongruent trials \((M = 675.59, SD = 103.93)\) were responded to significantly slower than congruent trials \((M = 631.45, SD = 97.10)\), \( t(20) = 2.67, p = 0.015 \). There was no significant difference in accuracy found between congruent \((M = 95.60\%, SD = 7.99)\) and incongruent trials \((M = 96.00\%, SD = 7.99)\). No differences between relatives of synesthetes and controls were found for the Stroop effect in RT or accuracy while in the scanner.

The Stroop effect and the variance in the Stroop effect (in RT) was larger in the scanner \((M = 44.14, SD = 75.87)\) compared to the computer lab \((M = 25.83, SD = 25.91)\), but this difference was not significant, \( t(20) = 1.13, p = 0.272 \). No significant difference was found in accuracy. The Stroop effect scores inside and outside of the scanner were not correlated, \( r(19) = .285, p = 0.210 \).
The Stroop effect and the amount of reading
A positive correlation between the amount of words read and the Stroop effect after reading would imply that the number of words read in color directly relates to the strength of the learned associations. In our previous study (Colizoli et al., 2012), we did not find such a correlation. The present study confirms this, as no correlation was found between the amount of words read and the post-reading Stroop effect.

Crowding task (outside of scanner)
Data for the crowding task is given in Table 5.3. If reading in color caused performance advantages based on the formation of letter-color associations at the perceptual level, we would expect to see an increase in accuracy on letters which had been colored in the book after reading compared to letters which had always been presented in black text (i.e. a 'pop-out' effect). Such a relationship would be indicated by an interaction between letter condition and session. However, no interaction was found between testing session and letter condition. In contrast, participants performed significantly better on the untrained letter condition compared to the trained letter condition in the pre reading session, t(21) = 2.134, p = 0.045 and post reading session, t(21) = 2.8, p = 0.011. This difference did not change between testing sessions. Overall participants performed the task above chance level at 25%.

No differences between relatives of synesthetes and controls were found for the crowding task either before or after reading.

<table>
<thead>
<tr>
<th>Pre Reading</th>
<th>Post Reading</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Letters</td>
<td>57.77 (14.56)</td>
<td>59.66 (13.97)</td>
</tr>
<tr>
<td>Trained Letters</td>
<td>52.97 (12.50)</td>
<td>53.28 (11.01)</td>
</tr>
<tr>
<td>Total</td>
<td>59.00 (16.88)</td>
<td>61.55 (16.46)</td>
</tr>
</tbody>
</table>

Crowding task (inside of scanner)
The log files of one participant was lost due to a technical error (control group), therefore, N = 21. While in the scanner, participants performed better on the untrained letter condition (M = 54.70%, SD = 18.43) compared to the trained letter condition (M = 43.25%, SD = 15.38) and this difference was significant, t(20) = 4.59, p < 0.001, in contrast to the expected effect.

No differences between relatives of synesthetes and controls were found for the crowding task in the scanner.

Performance on the crowding tasks dropped in the scanner (M = 48.97%, SD = 15.98) compared to in the computer lab (M = 61.48%, SD = 16.86), and this difference was significant, t(20) = 4.17, p < 0.001. However, participants performed the task in the scanner overall above chance level at 25%.
Preference for letter-color pairs

Before beginning the experiment, participants rated their preferences for 16 letter-color pairs using a 5-point Likert scale. Data are given in Table 5.4. There were no significant differences between the preference groups on the average preferred and non-preferred ratings. In order to test for an interaction between letter frequency and letter-color preference, a repeated measures ANOVA was carried out on the Stroop data in RT with the within-subjects factors of session (2 levels), congruency (2 levels, congruent and incongruent), and the between-subject factor of preference group (2 groups). The three-way interaction effect of session, congruency and preference group did not quite reach significance, \( F(1,20) = 3.51, p = 0.076, \eta^2 = 0.149 \). The between-subjects effect of preference group was not significant, \( F(1,20) = 2.36, p = 0.141, \eta^2 = 0.105 \). Post-hoc t-tests showed that there was a marginally significant difference between preference groups in the pre-reading Stroop effects, \( t(20) = 1.87, p = 0.076 \), but not the post-reading Stroop effects, \( t(20) = 0.55, p = 0.590 \) (Figure 5.3). The pre-reading Stroop effect in preference group 1 (\( M = 18.12, SD = 29.69 \); high frequency letters were given the most preferred colors) was larger than preference group 2 (\( M = -6.66, SD = 32.33 \)) and opposite in sign. The difference between pre-reading and post-reading Stroop effects was significant within preference group 2, \( t(9) = 2.69, p = 0.025 \), but not within preference group 1, \( t(11) = 0.48, p = 0.640 \). Group 1 showed a pre-existing Stroop effect while Group 2 did not (if anything it was reversed). This could be due to the fact that in Group 1, individuals received their most preferred colors for high frequency letters and their least preferred colors for the low frequency letters. Group 2 received their most preferred colors for the low frequency letters and their least preferred colors for the high frequency letters. It could be the case that in Group 1, these pre-existing preferences were stronger in high frequency letters exactly because they are more frequent. A pre-existing association with color may be stronger in more frequent letters, which may be evident in implicit and automatic letter-color associations measured with the Stroop task.

Independent-samples t-tests on several factors were conducted in order to determine if other variables could explain the difference found between preference groups. There were no significant differences between preference groups on age, number of books read, number of words read, consistency score, PA score, objective imagery score, while there was a marginal difference in VVIQ score, \( t(20) = 1.998, p = 0.060 \), such that preference group 1 scored higher, \( M = 2.51 (SD = 0.40) \), compared to preference group 2, \( M = 2.12 (SD = 0.51) \). In order to test whether VVIQ score was related to the change in the Stroop effect, we correlated the VVIQ score with the difference in the pre- and post-reading Stroop effects for each preference group. No significant correlations or trends were found. Therefore, it seems more likely that the difference in preference groups was due to an interaction between pre-existing letter-color pair preferences and letter frequency. An interesting line for future studies would be to investigate whether an individual’s preference itself will change in accordance with their newly formed associations.
Table 5.4 Preferences for letter-color pairs. Participants rated their preferences for letter-color pairs using a 5-point Likert scale before beginning the experiment. The questionnaire included 16 combinations of four letters presented in four colors (red, orange, green, blue). For preference group 1, we assigned their most preferred letter-color pairs to the high frequency letters (HF; letters: e, n), and their least preferred colored-letter pairs to the low frequency letters (LF; letters: a, r). For preference group 2, we assigned their most preferred pairs to the low frequency letters, and their least preferred pairs to the high frequency letters. Mean preferences and standard deviations are given.

<table>
<thead>
<tr>
<th></th>
<th>HF</th>
<th>LF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>4.21 (0.78)</td>
<td>2.08 (0.83)</td>
<td>3.15 (1.34)</td>
</tr>
<tr>
<td>Group 2</td>
<td>2.4 (0.75)</td>
<td>4.00 (0.65)</td>
<td>3.2 (1.07)</td>
</tr>
<tr>
<td>Total</td>
<td>3.39 (1.19)</td>
<td>2.95 (1.22)</td>
<td>3.17 (1.22)</td>
</tr>
</tbody>
</table>

Mean (St. Dev.)

Figure 5.3 Preference groups and the Stroop effect. A significant difference in the Stroop effect between testing sessions was found in preference group 2 (denoted with an asterisk) but not in preference group 1. Preference group 1 has a pre-existing Stroop effect, while preference group 2 does not. This graph illustrates that pre-existing preferences for letter-color pairs has an effect on the learned associations measured with the Stroop effect. Error bars represent the standard error of the mean.

Test of consistency
Participants were given a test-retest version of the synesthetic consistency test. The time in between test and retesting was on average 20 days (SD = 11.05). Overall consistency scores ranged from 9.30% to 55.81% (M = 32.35%, SD = 12.30; Table 5.5). Synesthetes typically score much higher (Eagleman et al., 2007), above 80% consistent (Baron-Cohen et al., 1987; 1993). Therefore, none of our participants were considered to be synesthetic when scoring the list as a whole. The subtests were also scored separately for letters (range = 7.69% to 53.84%, M =
28.50%, $SD = 12.37$), numbers (range = 10% to 100%, $M = 38.64\%, SD = 20.54$), and days of the week (range = 14.29% to 100%, $MS = 37.66\%, SD = 23.96$). One subject (participant 2) scored 100% on days of the week. Upon post-hoc questioning, she reported that the associations were conscious, but that she doubted whether they were ‘real’. Another subject (participant 8) 100% consistently matched different colors to the numbers on the number subtest, however, he reported upon post-hoc questioning that he did not have any conscious associations between numbers and colors. No other participants scored above 80% on any of the subtests.

There was no difference in the consistency scores between relatives of synesthetes and controls.

**Visual mental imagery**

Participants completed two questionnaires that subjectively and objectively tested the vividness of visual mental imagery. Data are given in Table 5.5. Note that the scale for the VVIQ is: $1 = $as vivid as a photograph, $5 = $no visual mental imagery). Scores on the VVIQ ranged from 1.34 to 2.94 ($M = 2.33, SD = 0.49$). Scores on the objective test of visual mental imagery ability ranged from 56% to 90% correct ($M = 79.14\%, SD = 8.45$).

There were no differences in either the subjective or objective visual mental imagery scores between relatives of synesthetes and controls.

**Projector-associator questionnaire**

After reading, participants completed the projector-associator (PA) questionnaire taken from Rouw and Scholte (2007). Data are given in Table 5.5. Half of the questions were ‘projector’ questions and the other half ‘associator’ questions. Possible scores on PA questionnaire ranged from -4 to 4. A score equal to zero reflects the fact that absolutely no synesthesia-like percepts were reported to be experienced, neither projector-like nor associator-like in nature. A positive score indicates a projector-type of synesthetic experience, while a negative score indicate an associator-type of synesthetic experience. Actual scores ranged from -1.67 to 0.17 ($M = -0.45, SD = 0.59$).

There was no difference in the projector-associator score between relatives of synesthetes and controls.
<table>
<thead>
<tr>
<th>Group</th>
<th>Preference</th>
<th>Age No. Books</th>
<th>Books Read</th>
<th>Word Count</th>
<th>Character Count</th>
<th>Stroop RT (ms)</th>
<th>Stroop % Consistency</th>
<th>PA VVIQ</th>
<th>Obj. Imagery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>27</td>
<td>25</td>
<td>17.76</td>
<td>85747</td>
<td>41672</td>
<td>12.82</td>
<td>2.50</td>
<td>76.92%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>30</td>
<td>30</td>
<td>19.57</td>
<td>85898</td>
<td>43579</td>
<td>15.78</td>
<td>2.94</td>
<td>79.49%</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>27</td>
<td>25</td>
<td>18.17</td>
<td>87125</td>
<td>45879</td>
<td>13.98</td>
<td>1.94</td>
<td>74.36%</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>30</td>
<td>30</td>
<td>18.57</td>
<td>88589</td>
<td>44579</td>
<td>12.82</td>
<td>2.50</td>
<td>76.92%</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>27</td>
<td>25</td>
<td>19.17</td>
<td>86689</td>
<td>43579</td>
<td>13.98</td>
<td>2.94</td>
<td>79.49%</td>
</tr>
</tbody>
</table>

Table 5.5: Individual differences data. Participants are ordered by the post-reading Stroop effect (in reaction times) from largest to smallest. Data for each participant is given concerning group identity (Syn. Group refers to relatives of synesthetes = 1, or matched controls = 2), age, amount of reading (character counts do not include spaces), Stroop effects (in reaction times = RT and accuracy = %, consistency score (total over all items), projector-associator score (PA), vividness of visual mental imagery score (VVIQ), and objective imagery score.

Reading (shortest counts do not include spaces’ Stroop effects (in reaction times = RT and accuracy = % consistency score (total over all items), projector-associator score (PA), vividness of visual mental imagery score (VVIQ), and objective imagery score. For each participant is given concerning group identity (Syn. Group refers to relatives of synesthetes = 1, or matched controls = 2). Age and objective imagery score.
Reading experience
After reading, participants completed a questionnaire designed to target their reading in color experience using a 5-pt. Likert scale (1 = do not agree, 5 = completely agree). The following questions were given in random order, but are presented here in logical order for convenience:

(Q1) I enjoy reading:  \( M = 4.45, SD = 0.67 \).
(Q2) I enjoyed reading in color:  \( M = 3.45, SD = 0.86 \).
(Q3) I enjoyed the content of the book:  \( M = 4.27, SD = 0.77 \).
(Q4) The colored text was aesthetically appealing (i.e. ‘pretty’):  \( M = 3.23, SD = 1.15 \).
(Q5) The colored text was ugly:  \( M = 1.82, SD = 1.18 \).
(Q6) The colored text was distracting:  \( M = 2.59, SD = 1.33 \).
(Q7) The colored text became less distracting over time:  \( M = 3.86, SD = 1.28 \).
(Q8) The colored text became more distracting over time:  \( M = 1.32, SD = 0.57 \).
(Q9) I felt as if I was reading faster in color (by the end of the book):  \( M = 2.09, SD = 1.27 \).
(Q10) I felt as if I was reading slower in color (by the end of the book):  \( M = 2.27, SD = 1.28 \).
(Q11) I was more motivated to read this book compared to a book with normal black text:  \( M = 2.55, SD = 1.26 \).
(Q12) I was less motivated to read this book compared to a book with normal black text:  \( M = 1.73, SD = 0.98 \).
(Q13) I like the color red:  \( M = 3.91, SD = 0.68 \).
(Q14) I like the color orange:  \( M = 3.23, SD = 1.15 \).
(Q15) I like the color green:  \( M = 4.00, SD = 0.76 \).
(Q16) I like the color blue:  \( M = 4.23, SD = 0.92 \).
(Q17) I read more than an average person:  \( M = 3.18, SD = 1.10 \).
(Q18) I read less than an average person:  \( M = 2.45, SD = 1.14 \).
(Q19) I tend to read books from the same genre:  \( M = 2.73, SD = 0.98 \).
(Q20) I tend to read books from a variety of genres:  \( M = 3.64, SD = 0.95 \).
(Q21) When I see certain letters in black text (e,n,a,r), I see them in color:  \( M = 1.05, SD = 0.21 \).
(Q22) When I see certain letters in black text (e,n,a,r), I experience them in color (i.e. in the mind’s eye):  \( M = 1.27, SD = 0.46 \).
(Q23) When I think about certain letters (e,n,a,r), I experience them in color (i.e. in the mind’s eye):  \( M = 1.59, SD = 1.10 \).
(Q24) When I think about certain letters (e,n,a,r), I see them in color:  \( M = 1.59, SD = 1.01 \).
(Q25) Whenever I see or think about letters, I have no color experience:  \( M = 4.23, SD = 0.97 \).
(Q26) Choose one of the following statements that best describes your experience… (A) When I see certain letters in black text (e,n,a,r), I see them in color: 0% chose this option. (B) When I see certain letters in black text (e,n,a,r), I experience them in color (i.e. in the mind’s eye): 0% chose this option. (C) When I think about certain letters (e,n,a,r), I experience them in color (i.e. in the mind’s eye): 21% chose this option. (D) When I think about certain letters (e,n,a,r), I see them in color: 18% chose this option. (E) Whenever I see or think about letters, I have no color experience: 61% chose this option.
(Q27) Have you noticed any changes in behavior or experience since you started reading the book(s)? Participants’ answers to this question were not constrained (i.e., they could write as
little or much as they pleased). Some interesting comments were: ‘...sometimes I tried to remember which letter had which color but this did not stick with me (#14).’ ‘I didn’t notice the colors of the letters after I had been reading for some days (#16).’ ‘I enjoy thinking of words that do not have AENR in them, so they are completely black (#19).’ ‘...overtime, I think I started not to notice anymore that the letters were colored (#13).’ ‘...it feels like I see broader. My (eye) focus seems to be wider (#1).’ ‘I found reading in color pleasant. It didn’t bother me at all and I found myself reading a little faster (#15).’ ‘I noticed that when flipping a page, the colors appeared as less distinctive as when I finished the previous page. I experienced it as some sort of grey blur with no particular clear color. More as 'shadow' letters. But if I continued reading, this sensation disappeared in (about) 5 sec. And then, the colors of the different letters appeared in distinctive clear colors again (#10).’

Does the self-reported experience of color predict the Stroop effect?
In our previous study (Colizoli et al., 2012), a positive correlation was found between the post-reading Stroop effect and the question ‘I am experiencing color when thinking about certain letters,’ suggesting that individuals who tended to agree more with this statement also had larger Stroop effects after reading. However, from this question, we could not differentiate whether the participants were indicating that they experienced color when thinking about the letters, seeing the letters, or in both cases. Therefore, in the present study, we have modified the questionnaire to be more specific concerning whether thinking or viewing letters may induce color experiences. The original question was replaced by two questions concerning experiencing color in the mind’s eye: Question 22, ‘When I see certain letters in black text (e,n,a,r), I experience them in color (i.e. in the mind’s eye)’ and Question 23, ‘When I think about certain letters in black text (e,n,a,r), I experience them in color (i.e. in the mind’s eye).’

As a confirmatory analysis, we tested whether we could replicate the original correlation by averaging the two questions into a single score and the averaged response ($M = 1.43, SD = 0.76$) was correlated with the post-reading Stroop effect. This correlation did not reach significance (1-tailed), $r(20) = .33, p = 0.067$. The correlation was however in the expected direction based on our previous study (Colizoli et al., 2012), suggesting that individuals who had larger Stroop effects were more likely to report experiencing color in the mind’s eye while seeing or thinking about the trained letters.

There were no differences found in this self-report rating between relatives of synesthetes and controls.

Are the questionnaires related to each other?
As an exploratory analysis, we correlated a subset of the questionnaire scores. We first tested the hypothesis that self-reported vividness of visual mental imagery is related to an objective measure of visual mental imagery abilities. No relationship was found between the VVIQ and objective imagery scores, $r(20) = 0.07, p = 0.748$, meaning that the objective and subjective ratings of visual mental imagery do not correspond to one another implying that people who rate their visual mental imagery as the most vivid do not necessarily score the highest on an objective test of visual imagery ability. Second, we tested whether the PA score (after reading) was related to both objective and subjective visual mental imagery. The variable PA was
marginally non-normal, therefore, we used Spearman rank-based correlations. We found that
the PA score significantly correlated with the objective imagery score, \( r_s(20) = -0.54, p = 0.010, \)
but not the subjective score (VVIQ), \( r_s(20) = 0.08, p = 0.737. \) This negative correlation between
PA and objective imagery scores means that the more an individual is classified as an
‘associator’, the more accurate they were on the objective test of visual mental imagery
abilities. It should be noted that the PA scores were almost entirely negative, meaning that no
one’s score was in the range of a projector-type synesthete (positive scores). After correcting
for multiple comparisons using the Bonferroni method, \( \alpha = 0.017, \) this correlation remained
significant.

There were no differences found in any of the questionnaire scores between relatives of
synesthetes and controls.

Do the questionnaire scores predict the Stroop effect?
As a final exploratory analysis, we tested whether the PA or VVIQ scores could predict the post-
reading Stroop effect. We did not correlate the Stroop effect with the objective imagery score,
because the PA score was significantly correlated with the objective imagery score. No
significant correlations were found, implying that the PA and mental imagery scores do not
predict the degree of learning the letter-color associations as measured by the Stroop test.

Neuroimaging results

Brain activation during the Stroop task
Participants completed the Stroop task while functional MRI scans were acquired. The scans of
one participant (relatives group) was lost due to a technical error, therefore, \( N = 21. \) In a whole-
brain analysis, we tested for differences in brain activation between congruent and incongruent
conditions as well as trained and untrained conditions. The trained condition consisted of both
the congruent and incongruent letter conditions, and the untrained condition consisted of the
untrained letter condition. Significant clusters of activation and local maxima for the contrasts
congruent > incongruent and incongruent > congruent are given in Table 5.6.

Activation for the congruent > incongruent contrast extended throughout much of the brain,
including bilateral activation in occipital-temporal, superior parietal and frontal regions,
including insular cortex (Table 5.6), which are areas found to be related to grapheme-color
synesthesia across several studies (Rouw et al., 2011). The peak voxel for the congruent >
incongruent contrast was in the left occipital lobe near areas V3 and V4 (xyz = -38, -90, -8). This
is interesting because in both conditions, participants viewed the letters in color, but congruent
trials activated the visual brain regions more than incongruent trials. The incongruent >
congruent contrast showed activated regions in the medial frontal pole, near the paracingulate
gyrus, as well as in the right middle temporal gyrus (Table 5.6). Frontal activation in this
contrast was expected as incongruent trials were harder than congruent trials and frontal brain
regions are known to be involved in cognitive control (Ridderinkhof et al., 2004). The role of the
middle temporal gyrus in synesthesia research is less well known. Van Leeuwen et al. (2010),
however, also reported activation in the middle temporal gyrus during a synesthetic priming
task in the incongruent > congruent contrast. This study was conducted on grapheme-color synestheses and this particular activation is in the same region as found in the current study.

Table 5.6 Effects of letter-color congruency in brain activation after reading in color during the Stroop task. Significant clusters and local maxima of fMRI activation related to congruent versus incongruent conditions during the Stroop task for the contrasts congruent > incongruent and incongruent > congruent. Whole brain Z-statistic values, MNI coordinates of the maximum Z-statistic (xyz), cluster size (voxels), and target brain regions are reported for each contrast of interest. Brain regions are based on the Harvard-Oxford Cortical Structural Atlas, the Juelich Histological Atlas, and Brodmann areas are reported from the Talairach Daemon when available.

<table>
<thead>
<tr>
<th>Congruent &gt; Incongruent</th>
<th>Cluster</th>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-max</th>
<th>Voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>L Inferior occipital gyrus (BA 18)</td>
<td>-38</td>
<td>-90</td>
<td>6</td>
<td>6.29</td>
<td>29537</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>R Inferior occipital gyrus (BA 18)</td>
<td>38</td>
<td>-90</td>
<td>7</td>
<td>5.94</td>
<td>52537</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Cerebellum</td>
<td>30</td>
<td>-52</td>
<td>24</td>
<td>5.98</td>
<td>29537</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>L Postcentral gyrus (BA 2)</td>
<td>-44</td>
<td>-34</td>
<td>26</td>
<td>5.00</td>
<td>29537</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>R Inferior parietal lobe, supramarginal gyrus (BA 40)</td>
<td>52</td>
<td>-30</td>
<td>50</td>
<td>4.3</td>
<td>29537</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>L Inferior parietal lobe, supramarginal gyrus (BA 40)</td>
<td>-50</td>
<td>-34</td>
<td>54</td>
<td>4.04</td>
<td>29537</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>L Superior parietal lobe (BA1/2)</td>
<td>-26</td>
<td>-44</td>
<td>72</td>
<td>3.85</td>
<td>29537</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>R Insular cortex (BA 13)</td>
<td>44</td>
<td>2</td>
<td>8</td>
<td>4.31</td>
<td>2778</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>R Inferior frontal gyrus (BA 44)</td>
<td>60</td>
<td>14</td>
<td>10</td>
<td>4.05</td>
<td>2778</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>R Superior frontal gyrus (BA 6)</td>
<td>14</td>
<td>44</td>
<td>46</td>
<td>3.32</td>
<td>995</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>mid. Cingulate gyrus (BA 32)</td>
<td>0</td>
<td>32</td>
<td>26</td>
<td>3.28</td>
<td>995</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incongruent &gt; Congruent</th>
<th>Cluster</th>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-max</th>
<th>Voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>R Paracingulate gyrus (BA 32)</td>
<td>2</td>
<td>44</td>
<td>16</td>
<td>3.53</td>
<td>711</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>L Paracingulate gyrus (BA 32)</td>
<td>-2</td>
<td>48</td>
<td>8</td>
<td>3.37</td>
<td>711</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>R Frontal pole (BA 10)</td>
<td>2</td>
<td>58</td>
<td>12</td>
<td>3.31</td>
<td>711</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>L Frontal pole (BA 10)</td>
<td>-2</td>
<td>68</td>
<td>12</td>
<td>3.4</td>
<td>711</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>R Middle temporal gyrus (BA 21)</td>
<td>54</td>
<td>-12</td>
<td>18</td>
<td>4.2</td>
<td>509</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>R Superior temporal gyrus (BA 21)</td>
<td>64</td>
<td>-12</td>
<td>6</td>
<td>3.31</td>
<td>509</td>
</tr>
</tbody>
</table>

Further investigation is necessary for understanding the role of the middle temporal gyrus in developmental and acquired forms of synesthesia.

For the contrasts trained > untrained and untrained > trained significant activation was found only in the untrained > trained contrast. All significant clusters of activation and local maxima for the contrast untrained > trained are given in Table 5.7. Activation in this contrast was widespread, seen in bilateral occipital lobe, the occipital-temporal junction, the parietal lobe, near the supramarginal gyrus and post-central gyrus, as well as the left putamen, left thalamus extending into the left insula.
**Table 5.7 Effects of trained versus untrained letters in brain activation after reading in color during the Stroop task.** Significant clusters and local maxima of fMRI activation related to trained (letters: e, n, a, r in congruent and incongruent colors) versus untrained letters (letters: b, g, k, t in all four colors) during the Stroop task for the contrast untrained > trained. No significant clusters were found for the contrast trained > untrained. Whole brain Z-statistic values, MNI coordinates of the maximum Z-statistic (xyz), cluster size (voxels), and brain regions are reported for each contrast of interest. Brain regions are based on the Harvard-Oxford Cortical Structural Atlas, the Juelich Histological Atlas, and Brodmann areas are reported from the Talairach Daemon when available.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-max Voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R Middle occipital gyrus (BA 18)</td>
<td>40</td>
<td>-88</td>
<td>6</td>
<td>5.69</td>
</tr>
<tr>
<td>1</td>
<td>R Inferior occipital gyrus (V4, BA 18)</td>
<td>42</td>
<td>-84</td>
<td>-10</td>
<td>5.48</td>
</tr>
<tr>
<td>1</td>
<td>L Inferior occipital gyrus (V4, BA 18)</td>
<td>-42</td>
<td>-88</td>
<td>-4</td>
<td>5.5</td>
</tr>
<tr>
<td>2</td>
<td>L Putamen</td>
<td>-26</td>
<td>-6</td>
<td>-2</td>
<td>3.34</td>
</tr>
<tr>
<td>2</td>
<td>L Insular cortex (BA 13)</td>
<td>-36</td>
<td>-4</td>
<td>14</td>
<td>3.31</td>
</tr>
<tr>
<td>3</td>
<td>L Thalamus</td>
<td>-16</td>
<td>-18</td>
<td>8</td>
<td>3.62</td>
</tr>
</tbody>
</table>

**Trained > Untrained**

**Untrained > Trained**

Differences in brain activation between relatives of synesthetes and matched controls during the Stroop task

Although no behavioral differences between relatives of synesthetes and controls were found on the Stroop task, differential brain activation between groups was still obtained. There were significant interactions (F-test) in brain activation between relatives and controls. Congruency effects and training related conditions were tested in separate whole-brain analyses (Table 5.8).

**Table 5.8 Interaction between relatives and controls in brain activation during the Stroop task.** Interactions between group (relatives of synesthetes vs. controls) and congruency (congruent vs. incongruent) and group and training (trained vs. untrained) were tested. Whole brain Z-statistic values, MNI coordinates of the maximum Z-statistic (xyz), cluster size (voxels), and brain regions are reported for each contrast of interest. Brain regions are based on the Harvard-Oxford Cortical Structural Atlas, the Juelich Histological Atlas, and Brodmann areas are reported from the Talairach Daemon when available.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-max Voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L Inferior occipital gyrus (V4, BA 18)</td>
<td>-38</td>
<td>-90</td>
<td>-18</td>
<td>3.62</td>
</tr>
<tr>
<td>2</td>
<td>L Postcentral gyrus (BA 1)</td>
<td>-58</td>
<td>-22</td>
<td>42</td>
<td>3.52</td>
</tr>
</tbody>
</table>

**Group X Congruency**

**Group X Training**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-max Voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R Inferior parietal lobe, angular gyrus (BA 40)</td>
<td>58</td>
<td>-56</td>
<td>40</td>
<td>3.25</td>
</tr>
</tbody>
</table>
The results of congruency showed that in the control group, congruent stimuli activated V4 and the postcentral gyrus (both in the left hemisphere) significantly more than incongruent stimuli compared to the relatives. Conversely, these same regions were significantly more activated in the brains of the relatives for incongruent compared to congruent stimuli.

The results of training showed that trained stimuli activated the right inferior parietal lobe along the angular gyrus significantly more than untrained stimuli in the control group, and untrained stimuli activated this same region significantly more than trained stimuli in the family-members group.

Most of the regions involved in these interactions are known to be related across studies of developmental synesthesia, for example, V4 in extrastriate cortex, the postcentral gyrus, and the inferior parietal lobe near the angular and supramarginal gyri (Rouw et al., 2011). It is interesting to see that although these groups do not differ behaviorally, there are still differences apparent in brain activation that does not seem to have a measurable effect on the behavioral data. These results are the first reported evidence of differential brain function between (non-synesthetic) relatives of synesthetes and controls during a letter-color Stroop task.

**Brain activation during the crowding task**

In a whole-brain analysis, the difference in brain activation during the crowding task between trained and untrained letter conditions was tested (note: these untrained letters were not the same as in the Stroop task and all stimuli were presented in black on a white background). The scans of one participant was lost due to a technical error (relative group), therefore, \( N = 21 \). The contrasts trained > untrained and untrained > trained were tested, but no significant activation was found for either contrast.

**Differences in brain activation between relatives of synesthetes and matched controls during the crowding task**

Although there were no significant differences found in the crowding task in the whole group \( (N = 21) \), differential brain activation between groups (relatives of synesthetes and controls) was still investigated as a whole-brain analysis. No significant clusters were found as the result of an F-test for an interaction between groups and training conditions. In an exploratory analysis, post-hoc contrasts revealed a trend towards an interaction between group and training conditions and are given in Table 5.9. The groups differed in activation when each contrasts was tested separately. Trained stimuli activated the hippocampus, medial frontal gyrus, inferior parietal lobe and dorsolateral prefrontal cortex more than untrained stimuli in the control group compared to the relatives. In contrast, these same regions were more activated in the brains of the relatives for untrained stimuli compared to trained stimuli. Still, no significant interaction between groups and training conditions was found.
Table 5.9 Brain activation of relatives and controls during the crowding task. Significant clusters and local maxima of fMRI activation (F-test) are given. Whole brain Z-statistic values, MNI coordinates of the maximum Z-statistic (xyz), cluster size (voxels), and brain regions are reported for each contrast of interest. Brain regions are based on the Harvard-Oxford Cortical Structural Atlas, the Juelich Histological Atlas, and Brodmann areas are reported from the Talairach Daemon when available. An F-test for the interaction between group and training conditions was not significant. Exploratory Post-hoc contrasts are reported below.

| Relatives > Controls |
| Untrained > Trained |
| Trained > Untrained |
| Cluster | Brain Region | x  | y  | z  | Z-max | Voxels |
| 1 | R Hippocampus | 6  | -40 | 0  | 3.87  | 1887   |
| 2 | R Medial frontal gyrus (BA 9) | 4  | 56 | 38 | 4.15  | 1049   |
| 3 | L Inferior parietal lobe (BA 39) | -42 | -66 | 48 | 4.00  | 812    |
| 4 | R Dorsolateral prefrontal cortex (BA 10) | 52 | 48 | -4 | 3.93  | 662    |
| 5 | L Insular cortex (BA 13) | -24 | 12 | 10 | 3.43  | 567    |

| Controls > Relatives |
| Untrained > Trained |
| Trained > Untrained |
| Cluster | Brain Region | x  | y  | z  | Z-max | Voxels |
| 1 | L Amygdala extending into hippocampus | -20 | 0  | -10 | 3.95  | 1889   |
| 2 | R Medial frontal gyrus (BA 9) | 4  | 52 | 42 | 4.48  | 1063   |
| 3 | L Inferior frontal lobe (BA 7) | -42 | -64 | 50 | 3.94  | 810    |
| 4 | R Dorsolateral prefrontal cortex (BA 10) | 52 | 48 | -6 | 3.64  | 619    |
| 5 | L Insular cortex (BA 13) | -28 | 12 | 10 | 3.63  | 563    |

(Synesthetic) color localization
In the post-reading MRI session, a color localizer was presented to participants. It was designed to localize color-responsive regions in the brain as well as to test for the presence of any ‘synesthetic’ color. Color regions were successfully localized, with peak activation found in the left temporal occipital fusiform gyrus (12,385 voxels, Z-max = 5.81, xyz = -34, -60, -18) and extending into bilateral V4 (Figure 5.4).

Additionally, we were interested to see if there was any effect of color when presenting the trained graphemes in black. If there is an effect of ‘synesthetic’ color, then we would expect to see a significant difference in the trained letters compared to untrained letters in areas that are known to process color information (all letters were presented in black and untrained letters in this localizer were different from the letters that were used for the untrained condition in the Stroop task and the untrained condition in the crowding task). Significant deactivation was found for trained letters compared to untrained letters (i.e. activation was greater when viewing untrained letters compared to trained letters; Figure 5.4) in the left middle occipital lobe near area V2 (4936 voxels, Z-max = 4.08, xyz = -26, -96, 10, Brodmann area 18) and in the
left parahippocampal gyrus on the border with the amygdala (599 voxels, $z = 3.63$, $xyz = -24$, -10, -22). The cluster of activation in the occipital lobe for the contrast $untrained > trained$ extended from the occipital pole into the lingual gyrus bilaterally, and included part of area V4.

These two contrasts were overlaid as a conjunction analysis in order to isolate common voxels significantly activated in both contrasts ($untrained > trained$ letters and $colored > untrained$ letters). Interestingly, the overlap between activation due to real and ‘synesthetic’ color was found in early visual areas bilaterally from V1 extending into V4 (Figure 5.4).

As an individual differences analysis, we tested whether we could predict the post-reading Stroop effect from average activation (Z-scores) related to the contrast $trained > untrained$ letters presented in black (Figure 5.4, in red; we note that the activation for this direction of the contrast was mostly negative). Activation for this contrast was averaged per participant within the region defined by the conjunction analysis (Figure 5.4, in yellow). A marginally significant negative correlation was found, $r(20) = -0.40$, $p = 0.068$, between brain activation ($trained > untrained$) and the post-reading Stroop effect (in RT), meaning that a larger Stroop effect was associated with more negative activation on this contrast (Figure 5.4).

The deactivation of trained letters compared to untrained letters in this localizer replicates the results found in the Stroop task (when the letters were presented in color). Importantly, this deactivation was found in voxels that were also sensitive to the presence of real color, implying that there may be an effect of color in the presence of these trained letters. It should be noted that there is a qualitative difference between these results and what has previously been found in developmental synesthetes. When significant ‘synesthetic color’ activation has been found in synesthetes, synesthesia-inducing stimuli (e.g., letters) are typically related to greater activation in extrastriate color areas, which sometimes include V4, compared to baseline (non-synesthesia inducing) stimuli or tasks (Rouw et al., 2011).

 Relatives of synesthetes and controls did not differ in activation related to real or ‘synesthetic’ color.
Figure 5.4 The neural basis of ('synesthetic') color. (A) The blocked design of the color localizer paradigm used that consisted of three letter conditions: trained letters presented in black, untrained letters presented in black, and untrained letters presented in color. (B) Brain activation related to veridical color (green), ‘synesthetic’ color (red) and their physical overlap in the brain (yellow). Masks of whole brain Z-statistic values are shown in MNI space. (C) The relationship between the Stroop effect and brain activation during the color localizer. The regions of overlap between veridical and ‘synesthetic’ color were used as a mask for the contrast trained > untrained letters and brain activation was marginally correlated with the post-reading Stroop effect. There were no differences between relatives and controls in the Stroop effect or brain activation related to color.
Experience-dependent plasticity in grey and white matter

We were interested to see whether experience-dependent plasticity in grey-matter and white-matter structure could be caused by reading in color. At the whole brain level, changes between testing sessions in grey-matter volume was investigated using VBM and white-matter coherence was tested with FA. No significant differences between testing sessions in either grey or white matter were found.

In order to narrow down the amount of voxels tested, post-hoc ROI analyses were conducted on grey matter using the significant functional activation related to training and congruency on the Stroop task as masks. Still, no significant changes in grey matter between testing sessions was found.

For white matter structure, the same ROIs were used for this analysis as described above (Rouw & Scholte, 2007). FA values in each sphere were averaged, and a paired-samples t-test was conducted on each ROI. Again, no significant differences between testing sessions in average white-matter values were found in any of the ROIs.

We cannot conclude that reading in color produced measurable changes in brain structure in this sample. Perhaps the amount of training was insufficient to induce structural brain changes or the analysis methods too insensitive to be able to measure small variations in brain structure.

Differences in grey matter between relatives and controls

Differences in grey-matter volume between relatives of synesthetes and controls was tested with voxel-based morphometry (VBM) in two ROIs defined by the significant interactions found in functional activation during the Stroop task (see Table 5.8) between (1) groups and congruency conditions (the two significant clusters found for this contrast were treated as one cluster for the current grey matter analysis; 1196 voxels), as well as between (2) groups and training conditions (458 voxels). The pre-training and post-training T1-weighted scans were combined for this group analysis.

No significant difference in grey-matter volume between groups was found for the congruency-ROI. Results of the training-ROI (located in right inferior parietal lobe in the angular gyrus) showed that controls had significantly more grey matter volume in a sub-region of this ROI compared to the relatives, $p = 0.020$ corrected FWE, cluster size = 50 voxels, xyz = 62, -48, 42. This result remains significant after Bonferroni correction over the two comparisons ($\alpha = 0.025$).

Differences in white matter between relatives and controls

Previously, increased structural connectivity related to synesthesia and measured with fractional anisotropy (FA) was found in four regions of the average white-matter skeleton (Rouw & Scholte, 2007). We used these four regions as ROIs defined a priori for the analysis of white-matter integrity. These regions were bilateral superior frontal lobe (beneath the central sulcus), left superior parietal cortex and right inferior temporal cortex near the fusiform gyrus. The pre-training and post-training diffusion-weighted scans were combined for this group
analysis. A spherical ROI was made (diameter = 15 mm) based on the coordinates reported by Rouw and Scholte (2007). FA values in each sphere were averaged. Thereafter, an independent-samples t-test was conducted on each ROI and corrected for multiple comparisons with the Bonferroni method (\(\alpha = 0.0125\)).

A marginally significant difference between relatives and controls in FA was found in the left frontal ROI \(t(20) = 2.62, p = 0.016\). Relatives had increased FA in this ROI \((M = 0.82, SD = 0.03)\) compared to controls \((M = 0.77, SD = 0.05)\). A trend towards significance was found in the same ROI in the right hemisphere, \(t(20) = 2.39, p = 0.027\), in which relatives \((M = 0.83, SD = 0.03)\) also had increased FA in this ROI than controls \((M = 0.79, SD = 0.04)\).

**Discussion**

We have successfully replicated our previous ‘reading in color’ study (Colizoli et al., 2012) in an independent sample of participants and across languages (English and Dutch). We have shown that automatic associations between letters and colors can be developed by simply reading books with consistently colored letters. In our previous study, we found a trend towards perceptual effects of ‘synesthetic’ color after training measured with the crowding task. In the current study, no such trend was found, indicating that the trained associations do not facilitate pop-out effects in the same way as veridical color (Hubbard et al., 2005). There was no evidence to suggest that reading in color affected low-level stimulus properties measured with the crowding task. It is worth noting that not all grapheme-color synesthetes showed effects of synesthetic pop-out on a similar task (Hubbard et al., 2005). Research suggests that improved detection of visual items found to be related to synesthesia is not in fact due to pre-attentive pop-out as previously believed, but related to the spatial location of the synesthetic color in reference to the veridical color (Ward et al., 2007; Ward et al., 2010). Therefore, any effects found to be related to synesthesia during the crowding task would provide support for low-level perceptual effects of color, but the presence of such effects could still not be used as a diagnostic marker of synesthesia.

The current study showed strong support of our previous finding that the degree to which the participants reported experiencing color when viewing and thinking about the trained letters correlated with the Stroop effect after reading. Although the correlation between the questionnaire score and the Stroop effect did not quite reach significance, it was in the expected direction. In addition, we were able to explore whether subjective and objective measures of visual mental imagery could also predict the post-reading Stroop effect. We found that not only were the subjective and objective measurements of visual mental imagery uncorrelated, neither was an accurate predictor of the strength of the learned associations. Therefore, the relationship between visual mental imagery and trained letter-color associations remains a topic of investigation.

In the current sample, we found that pre-existing preferences for letter-color pairs interacted with letter frequency in such a way that those individuals whose preferred letter-color pairs had been assigned to the low frequency letters (‘a’ and ‘r’) were driving the interaction between testing session and congruency. In other words, those individuals who received their
preferred letter-color pairs for the high frequency letters (‘e’ and ‘n’) already seemed to have a Stroop effect before reading began, and the size of this Stroop effect did not significantly increase after reading. These results indicate that implicit pre-existing letter-color associations can affect behavior in non-synesthetes, and furthermore, that these implicit associations may depend on their frequencies. Future research is needed in order to probe this relationship in addition to replicating the findings concerning letter-color pair preference.

The Stroop effect in brain activation
The trained letter-color associations were evident in differences in brain activity during the Stroop task when comparing congruent to incongruently colored letters. Activation during the Stroop task for the contrast congruent > incongruent was found in areas that are known to be involved in developmental synesthesia across multiple studies (Rouw et al., 2011), in regions such as the inferior occipital lobe near V4, inferior parietal lobe (near the angular gyrus), superior parietal lobe (along the supramarginal gyrus), precentral gyrus, and insular cortex. Interestingly, these same brain regions are found to be involved in synesthesia, while the synesthetic version of the Stroop task has not been employed in many fMRI studies on grapheme-color synesthesia (Rouw et al., 2011). Two studies have tested synesthetes using a comparable synesthetic Stroop task in an fMRI paradigm, but they did not report comparable contrasts (Specht & Laeng, 2011; Laeng et al., 2011). In the traditional Stroop literature, most attention has been given to the contrast of incongruent > congruent conditions, while relatively little attention has been given to the opposite contrast. Regarding the classic Stroop task, the contrast incongruent > congruent is typically associated with frontal lobe and anterior cingulate activation, which is thought to reflect cognitive control mechanisms. Cognitive control is necessary in order to inhibit an automatic response in the incongruently colored condition compared to the congruent condition, involving response and conflict monitoring (Leung et al., 2000). Our findings are in line with the frontal lobe activation typically found for this contrast. The contrast congruent > incongruent has received less attention however in the conflict processing literature (Roberts & Hall, 2008). The interpretation of brain activation related to trained letter-color associations in the Stroop task remains an interesting line for future research.

Brain activation related to training during the Stroop task
The trained letter-color associations were also evident in differences in brain activity during the Stroop task when comparing trained to untrained colored letters. Significant activation was not found for the contrast trained > untrained, but for the untrained > trained contrast in bilateral occipital lobe, the occipital-temporal junction, the parietal lobe, near the supramarginal gyrus and post-central gyrus, as well as the left putamen, left thalamus extending into the left insula. It is not uncommon that untrained stimuli show more activation than trained stimuli, perhaps due to habituation or a decrease in necessary attention in the presence of expected stimuli (Kassubek et al., 2001; Erickson et al., 2007). A similar pattern of results was found in a separate color localizer experiment, described below, in which the uncolored untrained letters evoked more activation than the uncolored trained letters (different sets of untrained letters were used between experiments).
Brain activation related to color localization

In addition to successfully localizing color-sensitive regions, a difference between trained and untrained letters presented in black was found in brain activation. Similarly to the Stroop task using colored letters, untrained letters presented in black were activated significantly more compared to trained letters. In a conjunction analysis, the size of the post-reading Stroop effect correlated with the difference in brain activation between the trained and untrained letter conditions presented in black typeface. This correlation was found in regions that were defined by the conjunction of real color and training-related brain activation. The correlation between this activation and the Stroop effect is strong evidence to suggest that the difference in brain activation upon viewing uncolored trained versus untrained letters in the visual cortex is directly related to learning the letter-color associations.

The brains of grapheme-color synesthetes have typically shown an increase in activation near area V4 when presented with inducing graphemes in black or grey typeface color compared to non-inducing graphemes (it is important to acknowledge that this type of ‘synesthesia’ contrast has not consistently shown significant synesthetic color activation across the literature; Rouw et al., 2011). In the current study, we found a decrease in ‘inducing’ graphemes (i.e. trained) compared to ‘non-inducing’ graphemes (i.e. untrained) in extrastriate cortex, including in area V4 (defined with a probability atlas). Similar to the results found during the Stroop task, where untrained letters also evoked more activation than trained letters (colored letters in that case), a possible explanation could be habituation of attention to expected stimuli (Kassubek et al., 2001; Erickson et al., 2007). Based on the correlation with the Stroop effect, an alternative explanation in this case could be that the relative deactivation was related to a ‘filtering’ process (perhaps mediated via higher-order brain areas) that reflected visual imagery of color in contrast to veridical perception (Amedi et al., 2005), or interpreted slightly differently, as an ‘error’ signal that reflected an expectation of color where there was none. As a follow-up to this result, we plan to evaluate activation related to the uncolored trained and untrained letters in retinotopically defined areas V1 through V4. We predict that if this deactivation is related to the experience of color (i.e. color constancy), then the correlation between brain activation and the Stroop effect should be significant in V4, but not V1 up to V3.

Differences between non-synesthetic relatives of synesthetes and matched controls

For the first time, we have shown that brain activation on the Stroop task after reading in color is different in non-synesthetic relatives of grapheme-color synesthetes compared to matched controls. The brain areas involved in the interactions between group and letter congruency and training included many of the areas related to developmental synesthesia, in particular related to the visual or perceptual aspect of synesthesia, such as the left inferior occipital gyrus in V4, and regions proposed to be related to the extraordinary ‘binding’ aspect of synesthesia, such as the right inferior parietal lobe along the angular gyrus. Relatives of synesthetes also showed differential brain activation during the perceptual crowding task, namely in the hippocampus, medial frontal gyrus, inferior parietal lobe, dorsolateral prefrontal cortex and insular cortex, even though no behavioral effects were significant.

There was no evidence for differences in behavior between the two groups on any task or questionnaire. Still, the brains of the relatives of synesthetes processed the stimuli differently
even though no behavioral measures could be differentiated based on group membership. The direction of the interaction between brain activation and Stroop congruency was unexpected: controls showed more activation than relatives during congruent compared to incongruent trials, in both the left postcentral gyrus and left V4. Controls also showed more activation than relatives during trained compared to untrained trials in the right angular gyrus. One explanation for the direction of these results might be that in these areas, the brains of the control group needed more activation compared to the relatives in order to achieve the same decision and motor output. In contrast, brain activation related to real or ‘synesthetic’ color did not differ between relatives and controls. Although the data are not reported here, no difference between groups was found for the visual word form area (VWFA) localizer. Both the color and VWFA localizers were passive tasks, in which the subjects did not have to respond unlike the Stroop and crowding tasks. Therefore, we speculate that perhaps these differences in brain activation would most likely be found during tasks that require a response from the participants.

A whole-brain VBM analysis of grey matter volume did not show any significant differences between relatives and controls. We performed a small volume analysis of grey matter in the regions defined by the significant interaction between group and training during the Stroop task. A region of the angular gyrus in the right inferior parietal lobe was activated differentially between relatives and controls when responding to trained versus untrained letters. Controls had significantly more underlying grey matter in a sub-region within this ROI compared to relatives. This difference in underlying brain structure may explain the difference in brain activation related to the trained letters. In line with the findings of brain structure in grapheme-color synesthesia (Rouw & Scholte, 2007), relatives showed evidence of increased FA in the left superior frontal lobe underneath the central sulcus (with a trend in the corresponding region in the opposite hemisphere). This is the first evidence to suggest that non-synesthetic relatives of synesthetes may also have increased structural connectivity in white-matter tracts. A complex pattern of brain structure and activation most likely underlie the difference between a synesthetic and non-synesthetic brain. The similarities between synesthetes and their relatives in brain structure may point to genetic pathways shared between synesthetic and non-synesthetic relatives, furthering the search for understanding the interaction between genes and environment in the development of synesthesia.

**Conclusion**

We conclude that training synesthetic associations by reading in color has an effect on brain activation in areas related to developmental synesthesia. Although there was no evidence for differential behavior in acquisition of synesthetic associations in relatives of synesthetes (as compared with controls), differences in brain activation due to learning were found in areas related to the sensory processing of the trained associations as well as areas involved in binding sensory information. Furthermore, underlying grey matter volume near the ‘binding’ area of the parietal lobe differed between relatives and controls. These results are consistent with the known literature on developmental synesthesia, while providing new insights into the interaction between genes and environment. Based on these results, it seems that although the brains of relatives of synesthetes are different from controls in areas related to synesthesia,
implying underlying genetic mechanisms, the presence of these genes is not enough to facilitate learning synesthetic associations compared to a closely matched control group. It remains to be seen whether more extensive training in non-synesthetic relatives of grapheme-color synesthetes can help to illuminate the environmental factors involved in the development of synesthesia.

**Acknowledgments**

We would like to extend a special thank you to both the publishers and authors at Nijgh & Van Ditmar (the Netherlands) for allowing us to use their books, as this research would not have been possible without the support of the publishers. We thank Joram van Driel for the experimental setup of the crowding task, Tomas Knapen and Renee Visser for their help with the EyeLink system, Nick Daems for writing the Visual Basic Word macro, Anja Pahor for her help testing participants in the MRI scanner, and Ilja Sligte for designing the retinotopic mapping paradigm. We would like to acknowledge the help of our synesthesia network for participant recruitment and especially thank all of our participants for their dedication.