Visual cortex activity predicts subjective experience after reading books with colored letters

Colizoli, O.; Murre, J.M.J.; Scholte, H.S.; van Es, D.M.; Knapen, T.; Rouw, R.

Visual cortex activity predicts subjective experience after reading books with colored letters

Olympia Colizoli [a,*], Jaap M.J. Murre [a], H. Steven Scholte [a,b], Daniel M. van Es [c], Tomas Knapen [c], Romke Rouw [a]

[a] Brain and Cognition, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
[b] Amsterdam Brain and Cognition Center, Institute for Interdisciplinary Studies, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[c] Cognitive Psychology, Vrije Universiteit, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

ABSTRACT

One of the most astonishing properties of synesthesia is that the evoked concurrent experiences are perceptual. Is it possible to acquire similar effects after learning cross-modal associations that resemble synesthetic mappings? In this study, we examine whether brain activation in early visual areas can be directly related to letter-color associations acquired by training. Non-synesthetes read specially prepared books with colored letters for several weeks and were scanned using functional magnetic resonance imaging. If the acquired letter-color associations were visual in nature, then brain activation in visual cortex while viewing the trained black letters (compared to untrained black letters) should predict the strength of the associations, the quality of the color experience, or the vividness of visual mental imagery. Results showed that training-related activation of area V4 was correlated with differences in reported subjective color experience. Trainees who were classified as having stronger 'associator' types of color experiences also had more negative activation for trained compared to untrained achromatic letters in area V4. In contrast, the strength of the acquired associations (measured as the Stroop effect) was not reliably reflected in visual cortex activity. The reported vividness of visual mental imagery was related to veridical color activation in early visual cortex, but not to the acquired color associations. We show for the first time that subjective experience related to a synesthesia-training paradigm was reflected in visual brain activation.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The prevalence of culturally dependent inducers (e.g. language, numbers, time) of synesthetic experiences begs the questions to what extent synesthetic associations are 'picked up' directly from the environment and whether synesthesia can be learned given enough training. Associative learning has long been speculated to play a role in the mechanisms behind the development of synesthesia (Howells, 1944; Kelly, 1934). Developmental synesthesia might arise because individuals with a certain genetic predisposition have a relatively low threshold for forming hard-wired connections between cross-modal and cross-sensory stimuli that tend to co-occur in the environment. However, the mechanism of developmental synesthesia does not necessarily preclude the possibility that non-synesthetes can acquire the same behavior

and experience as synesthetes from high-frequency exposure to cross-modal stimuli. Bor et al. (2014) had participants complete an intensive nine-week synesthesia-training paradigm that consisted of several different tasks, some of which were adaptive to the level of the individual's performance. Arguably, the most striking characteristic exhibited by the participants in this study was the reported phenomenological experience of color in the absence of the corresponding veridical color (although this did not persist for the participants three months after training ceased). Such results imply that our experience of the world is more modifiable than previously believed. It is still unknown how acquired phenomenological associations affect the brain. In the current study, we investigate the relationship between letter-color associations acquired by reading books with colored letters and brain function in the visual cortices of non-synesthetes while the participants viewed achromatic and colored letters.

Letters are presented thousands of times in a typical novel, which are normally printed in black text. Reading books with

colored letters is therefore a feasible associative training paradigm; it is primarily an incidental form of learning, although the participants are aware of the colored letters (Colizoli et al., 2012, 2014a). While the focus when reading is on the meaning of the text, repeatedly seeing the same letters in the same colors has an effect on behavior. We have previously shown that automatic associations between letters and colors could be acquired by simply reading books with consistently colored letters, as measured by the 'synesthetic' Stroop effect (Colizoli et al., 2012). We stress that the Stroop effect found after reading the colored books reflects the automatic, but not necessarily perceptual, component of the acquired associations between letters and colors (MacLeod and Dunbar, 1988). In addition, a positive correlation was found between participants answers to the question '*I am experiencing color when thinking about certain letters*' and the size of the Stroop effect after reading, indicating that individuals who tended to agree with this question more also showed greater Stroop effects. In contrast, no correlation was found between answers to the question '*I am experiencing color when I see certain letters*' and the size of the Stroop effect after reading. Although we could not discern from these two questions alone exactly what these participants were experiencing or thinking about, the results led us to hypothesize that individual differences in visual mental imagery or the quality of the subjective experience (i.e. the location or intensity of a mental color experience) would be factors related to developing an acquired Stroop effect. In addition, we predicted that evidence of color (perhaps as color imagery or the expectation of color) would be apparent in the visual cortices of non-synesthetes while viewing the achromatic trained stimuli after they had completed a synesthesia-training paradigm.

Mental imagery and synesthetic experiences seem to be closely related but are not the same. For example, in a well-controlled color imagery experiment, Rich et al. (2006) found that color imagery activated a region near area V4 in both synesthetes and controls, while synesthetic color activated the left medial lingual gyrus for the synesthetes only. Synesthesia is furthermore qualitatively and quantitatively different from spontaneous mental imagery, because the experienced associations are automatically and reliably evoked, highly specific and consistent over long periods of time, and are seemingly arbitrary while having substantial bandwidth (i.e. the number of mappings) (Colizoli et al., 2014b; Rothen and Meier, 2014; Mylopoulos and Ro, 2013). Interestingly, synesthetes reported having more vivid visual mental imagery (not necessarily related to their synesthesia) compared to controls (Barnett and Newell, 2008) as measured with the vividness of visual mental imagery questionnaire (VVIQ) (Marks, 1973). The VVIQ scale ranges from '*no visual experience at all (only knowing that something is the case)*' to having a mental experience that is '*as clear and vivid as normal vision.*' Similarly, the diversity of the reported experiences of the synesthetic concurrent sense is an indication that the phenomenology for some individuals is more perceptual in nature (i.e. as clear and vivid as normal vision) than for others (Simner, 2013). Therefore, we hypothesized that those individuals who report experiencing more vivid mental imagery would be more likely to acquire letter-color associations.

In addition to being more or less vivid, synesthetic experiences can have other qualities. For example, the experiences can or cannot be localized in physical and mental space. Some synesthetes locate the synesthetic color outside of themselves (i.e. projected into the world), while others describe the color experience as being 'in the mind's eye.' Dixon et al. (2004) termed this the 'projector-associator' (PA) distinction. The PA distinction has been a fruitful approach in brain imaging research on synesthesia. The PA distinction of synesthesia has been related to specific neural markers (Terhune et al., 2015; Rouw and Scholte, 2010, 2007) and brain dynamics (van Leeuwen et al., 2010, 2011).

Intuitively, it seems that the projector-type of synesthetic experiences more closely resemble normal vision than the associator-type. We would therefore expect that forms of newly acquired synesthetic experiences via associative training would fall in the associator-range of experiences and not in the projector-range for most individuals. Furthermore, we examined if the quality of the color experience measured by the PA dimension might be evident in brain activation within the visual cortices of non-synesthetes participating in a synesthesia-training paradigm.

In the current study, participants read specially prepared books with colored letters for several weeks and were subsequently scanned using functional magnetic resonance imaging (fMRI) in order to assess the effects of repeated letter-color mappings (i.e. training) on brain function. A standard retinotopic mapping technique (Wandell et al., 2007) and a visual word form area localizer (VWFA) (Baker et al., 2007) were applied before training began in order to independently assess the contributions of the different regions of interest (ROIs): areas V1, V2, V3, V4 and the VWFA. A color localizer was administered after training was completed to test for activation related to training effects and veridical color. The acquisition of the letter-color associations was verified using a synesthetic version of the classic Stroop task (Colizoli et al., 2012, 2014a). Vividness of visual mental imagery was assessed using the VVIQ (Cui et al., 2007; Marks, 1973). The projector-associator questionnaire (Rouw and Scholte, 2010, 2007) was administered in order to assess the nature of any (possible) color experiences in the absence of veridical color after training. At the behavioral level, we expected that individuals who report experiencing more vivid visual mental imagery or more color experiences would also be the most susceptible to acquiring letter-color associations (measured as a stronger Stroop effect). If the acquired letter-color associations are visual in nature (even as a visual mental image), then brain activation in visual cortex while viewing the achromatic trained letters (compared to achromatic untrained letters) are expected to predict the strength of the associations, the quality of the color experience, or the vividness of visual mental imagery. We furthermore investigated whether relative brain activation in visual cortex while viewing untrained letters presented in veridical color (compared to achromatic letters) was related to any of the behavioral variables of interest.

2. Materials and methods

2.1. Participants

Twenty-two subjects participated in the study (16 women, $M = 24.95$ years, $SD = 2.44$, range $= [22, 30]$ years). Half of the sample was a (1st, 2nd or 3rd degree) relative of a grapheme-color synesthete; the other half was matched on age, gender, handedness and education. For the current study, we evaluated all participants as a single group ($N = 22$) for the individual differences analysis after verifying that none of the variables of interest differed by group. Participants were screened by questionnaire and interview with the researcher for synesthesia, dyslexia, and attention-deficit disorder (ADD), in addition to standard MRI-related risk factors. Participants were tested for color blindness using the Ishihara test for color blindness (Ishihara, 1936). The Ethical Committee of the Department of Psychology at the University of Amsterdam approved this experiment. 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. No MRI runs were discarded due to motion (< 1 mm in all directions) or other artifacts. One run of one participant (relative group) for the Stroop task in the scanner was lost due to

technical error. The behavioral log files of one participant (non-relative group) for two runs of the Stroop task in the scanner were lost due to technical error.

2.2. Experimental procedure

We aimed to replicate the behavioral results of our previous study in a different sample (Colizoli et al., 2012). Therefore, we decided not to make major changes to the experimental paradigm, for example, by introducing new (control) conditions. Our paradigm did not allow us to fully balance the letter conditions between or within individuals, because we required that the consistently colored letters in the books were viewed as often as possible; therefore, we chose very frequent letters to be presented in color. Before testing began, we determined letter-color pairs separately for each participant using a procedure identical to the one reported in Colizoli et al. (2014a). All participants read specially prepared books that contained four high-frequency lower-case letters ('a', 'e', 'n', and 'r') in four high-frequency colors (red, orange, green and blue). Participants were instructed to always read the books as they normally would but always under good lighting conditions to ensure the colors were visible. They were instructed to keep track of how much they read (by marking in the books where they stopped before the second testing session in case it was not at the end of the book) and to inform the researcher before they ran out of reading material so that a new book could be delivered. We emphasized honesty in reporting how much was read and explained that it was crucial to get an accurate representation of the amount of reading for the research goals. 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 content of the books was not altered in any way. The formatting of the colored books for reading can be found in Colizoli et al. (2014a).

All participants were tested before and after training. Before and after reading, participants completed both a behavioral and MRI testing session and filled in several questionnaires. The order of the behavioral and MRI sessions were counterbalanced between participants, but remained the same for each individual when possible. In the behavioral sessions, participants completed a Stroop task, a crowding task and were asked to read text on the screen while eye movements were recorded (crowding and reading data not reported here). The order of the behavioral tasks was counterbalanced between participants but remained the same for each individual across testing sessions. Each behavioral session took approximately one hour to complete. Participants were seated 50 cm in front of a 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 1280×1024 pixels. All responses were recorded with a USB keyboard. In the pre-training testing session, participants completed a general screening form, the first part of a test of consistency, and the VVIQ (Cui et al., 2007; Marks, 1973). In the post-training testing session, participants completed the second part of the consistency test, the PA Questionnaire (Rouw and Scholte, 2010, 2007), and a general questionnaire about their reading experience (Colizoli et al., 2014a).

For the MRI procedure, the functional tasks differed between testing sessions. Functional (T2*-weighted) and structural data (T1-weighted and diffusion-weighted) 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). During the pre-training MRI session, a VWFA localizer and a retinotopic mapper were administered. The VWFA localizer always preceded the retinotopic-mapping runs. During the post-training MRI

session, the Stroop task and crowding task (data not reported here) were administered in addition to a color localizer. The color localizer always preceded the Stroop and crowding tasks. All participants were monitored with an eye tracker to ensure that no one fell asleep. Each MRI session took 1.5 h of scanning time.

Subjects were debriefed about color phenomenology at the end of the post-training testing session with two questionnaires: the 'reading experience questionnaire' that had been adjusted from our earlier study (Colizoli et al., 2012; 2014a) and the PA questionnaire that had previously only been given to synesthetes (Rouw and Schote, 2007; 2010). Both of these questionnaires asked about conscious color phenomenology upon viewing or thinking about black letters. Participants were able to ask clarification questions to the researcher at any time. There were no specific questions aimed at their experiences during the color localizer. Participants were given space to openly report any other comments.

We do not report the data on the crowding task (Colizoli et al., 2012) for the following reasons. An untrained letter condition was intended to serve as a within-subjects baseline measure for the crowding task. In the absence of the predicted effect of enhanced performance on trained letters after training, we should at least see equal performance between the two letter conditions before training to assure a proper baseline measure. Based on the fact that the untrained letters were identified significantly better than the trained letters in the behavioral and MRI sessions, and this did not change related to reading, we can conclude that the crowding task was not effective. The difference found between letter conditions must be attributed to the differences in the physical letters that make them more or less easy to 'crowd' and therefore meaningless for our purposes here. We note that in the previous study, uppercase letters were trained in the books and used in the crowding task. Here, we trained only lowercase letters and these were used in the crowding task.

2.3. Test of consistency

We aimed to objectively verify that none of our participants had grapheme-color synesthesia using a test of consistency. We followed the questionnaire test-retest paradigm (Asher et al., 2006; Baron-Cohen et al., 1987). The cutoff for considering a 'diagnosis' of synesthesia was greater than 70% consistent across the 43 items (Asher et al., 2006; Baron-Cohen et al., 1987). During the pre-training testing session, participants were presented with a list of 26 letters (all uppercase), 10 digits (0–9) and seven days of the week (43 items in total). Participants were asked: if each letter, number or day were to have a color, what color would it be? They were instructed to indicate the color of each item by filling in the questionnaire. They were instructed not to list the same color for all items (e.g. as all 'black'). The participants were not informed that they would receive this questionnaire again. During the post-training testing session, they were again given the same list of 43 items again in randomized order. The instructions were the same as during the pre-training session.

2.4. Stroop task

The procedure for the Stroop task followed previously described methods (Colizoli et al., 2014a) with the following exceptions: Participants were shown one of eight letters ('a', 'b', 'e', 'g', 'k', 'n', 'r', or 't') in one of four colors (red [RGB=230, 0, 0], orange [255, 143, 0], green [0, 181, 0] or blue [0, 125, 255]). The letter-color pairs comprised three conditions: (1) the *congruent* condition consisted of four trained letters ('a', 'e', 'n', and 'r') presented in colors congruent with the colors presented for each individual within the books, (2) the *incongruent* condition consisted of four

trained letters ('a', 'e', 'n', and 'r') presented in colors incongruent with the colors presented for each individual within the books, and (3) the *neutral* condition consisted of four untrained letters ('g', 't', 'k', and 'b', that were always in black text within the books) presented in the four colors used during training. The *congruent*, *incongruent* and *neutral* trials were randomized, with a total of 288 trials (96 *congruent*, 96 *incongruent*, and 96 *neutral* trials). Reaction times that were greater than 2.5 times the standard deviation per participant and condition were removed. We expected to replicate our previous result of the two-way ANOVA interaction between testing session (pre-training vs. post-training) and congruency (*congruent* vs. *incongruent*) (Colizoli et al., 2012). The *neutral* letter condition was used as an exploratory measure. Participants were instructed to indicate (via a keyboard press) the color of the letter presented as fast and accurately as possible.

2.5. Color localizer

A color localizer was used to test for voxels that respond to the achromatic trained stimuli and veridical color modeled after van Leeuwen et al. (2010). It consisted of three stimulus 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 *colored* letters ('c', 'm', 'v', and 's') presented in eight distinct colors (red [RGB=230, 0, 0], orange [255, 102, 0], brown [179, 110, 0], yellow [246, 236, 0], green [0, 210, 0], blue [0, 203, 230], purple [150, 45, 225] and pink [242, 112, 220]). Our rationale behind the stimuli used for the color localizer was the following. As this experiment was the first of its kind, we had no information concerning whether any training-related differences would be obtained using this specific paradigm. Therefore, we aimed to keep the context of the letters in all conditions of the color localizer as visually similar as possible to the way that the letters appeared in the books. Within the books, all letters were always presented in the font 'Arial Black' and all *untrained* letters were presented in black text against a white background. Therefore, in the localizer, we used the same font (Arial Black) as presented in the books and also chose to present the letters in black font color for both the *trained* and *untrained* (i.e. achromatic) letter conditions. This prevented us from greyscale-matching the luminance of the colored letters to the luminance of the achromatic letters within the color localizer. Therefore, the contrasts of veridical color contain effects of luminance differences in addition to differences between hues. The background during the color localizer was light gray [RGB=217, 217, 217]. We did not use a white background, because during piloting it was reported that there was insufficient contrast on the screen in the scanner when the letters were presented in yellow against a white background. Changing the background towards gray was reported to have helped with the contrast of yellow letters against the background. For the *colored* condition, we chose to use different variations of hues of red, orange, green and blue compared to those that were trained, as we aimed not to introduce possible confounds related to training in the veridical color condition.

The color localizer was a blocked design with 16-s stimuli blocks and 16-s periods of rest between blocks (fixation cross). Conditions were presented in pseudo-randomized blocks and each condition was presented six times (18 blocks in total). Within each block, 16 stimuli were randomly presented (four letters with four repetitions) for 500 ms with an inter-trial interval of 500 ms. For each block of condition 3, each of the eight colors was shown twice and the order of the colors remained constant, while the four letters were randomly assigned (with four repetitions) to each color. The color localizer consisted of 330 volumes and lasted 11 min. Participants passively viewed the stimuli. The instructions were to focus on the middle of the screen during the duration of the localizer, both when the letters appeared and also when the fixation cross was presented.

2.6. Visual word form area (VWFA) localizer

A VWFA localizer was used to define the VWFA for each participant (Qiao et al., 2010; Baker et al., 2007; Dehaene et al., 2004; Polk and Farah, 2002; Dehaene et al., 2001). The VWFA localizer was a blocked design with 16-s stimuli blocks and 16-s periods of rest between blocks (fixation cross). It consisted of two stimulus conditions: (1) Dutch words with four letters and equal average frequencies chosen at random, and (2) The same words from condition 1 presented as Chinese characters in Hanzi Kaishu font. Each condition was presented nine times (18 blocks in total) in alternating order (i.e. ABAB). Within each block, 16 words were randomly drawn from a list of 144 words per condition without repetition and presented for 500 ms with an inter-trial interval of 500 ms. All words were presented in black Courier New font against a white background. The VWFA localizer consisted of 330 volumes and lasted 11 min. Participants passively viewed the stimuli. The instructions were to focus on the middle of the screen during the duration of the localizer, both when the words appeared and also when the fixation cross was presented.

2.7. Retinotopic mapping

Retinotopic visual areas were localized following standard protocol (Wandell et al., 2007). The retinotopic mapper consisted of two polar-angle mapping (wedge stimulus) runs and one eccentricity mapping (expanding ring) run. Each functional run of the retinotopic mapper (three runs in total) consisted of 270 volumes and lasted 4.5 min. During all retinotopic-mapping runs, participants indicated a change in the color of the fixation cross with a key press. Fourier-based techniques were used to determine the preferred polar phase and *F*-test statistic of each voxel in visual cortex. Polar phase and eccentricity preference values were color-coded and projected onto an inflated structural T1-weighted image for each subject (Wandell et al., 2007). Borders between the separate visual areas were drawn manually for areas V1, V2v, V2d, V3v, V3d, V3AB, and V4 according to standard criteria (Wandell et al., 2007).

2.8. Image acquisition

Scans were acquired on a Philips 3T Achieva TX scanner, located at the Spinoza Center, Amsterdam, the Netherlands. Whole brain gradient-echo echo-planar imaging (EPI) measurements (voxel size=$3 \times 3 \times 3$ mm^3, repetition time [TR]=2000 ms, echo time [TE]=27.63 ms, flip angle=76.1°, FOV=240×240, matrix=80×80, 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. Scans for the retinotopic mapping runs consisted of partial volume scans (voxel size=$2.5 \times 2.5 \times 2.5$ mm^3, repetition time [TR]=2000 ms, echo time [TE]=34.53 ms, flip angle [FA]=76.1°, FOV=200×200, matrix=80×80, slice thickness=2.5 mm, slice gap=0.25 mm, 24 slices per volume, sensitivity encoding factor of 2). Two T1-weighted anatomical scans were acquired per session (four T1 volumes per participant, voxel size=$1 \times 1 \times 1$ mm^3, TR=8229 ms, TE=3.77 ms, flip angle=8°, FOV=256×256, matrix=256×256, slice thickness=1 mm, no slice gap, 160 slices per volume). Only one T1-weighted image was used for registration purposes.

Table 1

Individual differences data for behavioral measures and questionnaires. The table is ordered by the projector-associator (PA) score for convenience. Data for each participant is given concerning group membership (relative of a synesthete=1, non-relative=2), amount of reading (word and character counts, spaces not included), consistency scores (percentage of a total of 43 items), Stroop effects for the post-training reaction time (RT) data (in milliseconds), vividness of visual mental imagery scores (VVIQ), projector-associator (PA) scores including the associator (Assoc.) and projector (Proj.) scores, and responses to question 30 (Reading Exp. Q30) from the reading experience questionnaire (Colizoli et al., 2014 a). The possible responses for Q30 were: (a) 'When I see certain letters in black text (e, n, a, r), I see them in color.' (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).' (c) 'When I think about certain letters (e, n, a, r), I experience them in color (i.e. in the mind's eye).' (d) 'When I think about certain letters (e, n, a, r), I see them in color.' (e) 'Whenever I see or think about letters, I have no color experience.' Mean (M) and standard deviation (SD) are given for the numerical data.

Group	Word count	Character count	Consistency (%)	Stroop RT	VVIQ	PA	Assoc.	Proj.	Reading exp. Q30
2	88,986	433,757	16.28	63.78	2.81	−1.67	2.67	1.00	c
1	34,982	171,405	55.81	39.38	1.34	−1.50	3.17	1.67	c
2	56,807	283,656	30.23	41.65	2.59	−1.33	3.17	1.83	d
1	58,065	266,666	34.88	22.36	2.75	−1.33	3.00	1.67	d
2	73,125	338,759	30.23	18.93	2.66	−1.17	2.17	1.00	c,e
2	91,212	423,076	13.95	4.30	2.00	−0.83	1.83	1.00	c
1	33,164	153,301	23.26	24.70	1.72	−0.67	1.67	1.00	e
2	180,423	835,654	34.88	43.88	2.41	−0.50	2.83	2.33	d
2	95,175	438,967	30.23	27.27	2.13	−0.50	1.50	1.00	c
2	43,714	209,931	37.21	1.92	1.75	−0.33	1.33	1.00	e
1	75,889	380,750	30.23	59.30	2.06	−0.17	1.33	1.17	d
1	58,065	266,666	16.28	41.73	1.63	0.00	1.00	1.00	e
1	58,065	266,666	39.53	32.12	2.88	0.00	1.00	1.00	e
1	68,353	322,812	16.28	21.32	1.94	0.00	1.00	1.00	e
1	46,574	231,568	9.30	18.20	2.78	0.00	1.00	1.00	e
1	165,178	818,795	39.53	13.71	2.75	0.00	1.00	1.00	e
2	101,084	468,530	34.88	13.83	1.91	0.00	1.00	1.00	e
2	118,043	554,853	44.19	61.40	2.50	0.00	1.00	1.00	e
1	68,353	322,812	39.53	52.97	2.06	0.00	1.00	1.00	e
2	75,061	356,033	39.53	−31.90	2.94	0.00	1.00	1.00	e
2	58,065	266,666	51.16	32.01	2.67	0.00	1.00	1.00	e
1	120,362	570,853	44.19	−34.53	2.94	0.17	1.00	1.17	e
M:	80,397.50	381,008.00	32.35	25.83	2.33	−0.45	1.62	1.17	–
SD:	38,278.92	182,383.01	12.30	25.91	0.49	0.59	0.82	0.36	–

2.9. fMRI preprocessing and statistical analysis

The fMRI images for the color localizer and VWFA localizer were analyzed using the FMRI Expert Analysis Tool (FEAT) version 6.00, part of University of Oxford's Functional MRI of the Brain Software Library (FSL) version 5.0.4 (http://www.fmrib.ox.ac.uk/fsl). Preprocessing steps included motion correction (Jenkinson et al., 2002), spatial smoothing using a Gaussian kernel of full-width at half-maximum (3 mm), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high-pass temporal filtering using a Gaussian-weighted least-squares straight line fitting ($\sigma=48$ s). Voxels belonging to brain tissue were extracted from non-brain tissue voxels using the Brain Extraction Tool (BET) (Smith, 2002). 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 $p=0.05$ for the color localizer and $p=0.001$ for the VWFA localizer following Yeatman et al. (2013). Resulting contrast images were linearly registered to the anatomical structure using FMRIB's Linear Registration Tool (FLIRT) with 7° of freedom and the full search space (Greve and Fischl, 2009; Jenkinson and Smith, 2001), then spatially normalized to the T1-weighted MNI-152 stereotaxic space template (2 mm) using FMRIB's Non-Linear Registration Tool (FNIRT) with 12 degrees of freedom and the full search space. The higher-level analyses were carried out using an analysis of mixed effects (FLAME) stages 1 and 2 with automatic outlier detection. Z-statistic (Gaussianised T/F) images were thresholded using clusters determined by $Z>3$ and a corrected cluster significance threshold of $p=0.05$.

2.10. Regions of interest selection

ROIs were determined a priori as visual areas V1, V2, V3, V4 and the VWFA. All ROIs were defined in each subject's native space.

Each retinotopic region was transformed using FLIRT from surface space into volumetric space based on the subject-specific transformation matrices determined in the first level analysis of the color localizer. Thereafter, dorsal and ventral components for areas V2 and V3 were combined. The two hemispheres for areas V1, V2, V3 and V4 were combined. A binary mask for each retinotopic area was created for each participant. The VWFA was defined per participant (in volumetric space) as all significant voxels (uncorrected at $p<0.001$) for the contrast of *Dutch > Chinese* words in the ventral occipital-temporal cortex of the left hemisphere only (Yeatman et al., 2013). A binary mask of the VWFA was created from the resulting voxels and transformed to the native space of each participant for the color localizer using FLIRT. All overlapping voxels between the ROIs were removed for the analyses. The uncorrected activation (Z-stat) for the contrasts of interest of the color localizer was averaged across all voxels per ROI per participant.

3. Results

The significance level for the repeated measures ANOVAs and post hoc t-tests conducted was $\alpha=0.05$. A total of 18 variables were investigated in the individual-differences correlation analysis, resulting in a correlation matrix of 324 possible comparisons. The false discovery rate (FDR) (Benjamini and Hochberg, 1995) was used to correct for multiple comparisons that included all reported behavioral correlations and brain-behavior/experience correlations. The FDR-adjusted significance threshold was $\alpha=0.016$. All correlations reported are non-parametric rank-based correlations and are two-tailed unless otherwise specified.

3.1. Behavioral results

All reaction time data are reported in milliseconds (ms) and accuracy data reported are the percentage of correct responses.

3.1.1 Test of consistency

Results for the test of consistency are presented in Table 1. The time between the first and second version of the questionnaire was on average 20 days ($SD=11$). Consistency scores ranged from 9.30% to 55.81%. None of the participants were considered to have synesthesia based on the test of consistency for the 43 items, because all scores were below 70% consistent. Synesthetes typically score above this threshold over longer periods of time than were tested here (Asher et al., 2006; Baron-Cohen et al., 1987). Importantly, none of the subjects reported having any conscious perceptual grapheme-color associations (Colizoli et al., 2014b).

3.1.2 Acquired letter-color associations from reading in color

We aimed to confirm that participants learned the associations from reading the colored books to the point that they interfered with task demands at the level of the whole sample tested. The strength of the learned letter-color associations acquired by reading the colored books was tested with a 'synesthetic' version of the classic Stroop task following our reading-in-color training paradigm outlined in Colizoli et al. (2014a). Participants read the colored books within an average of 20 days ($SD=11$). The number of words and characters read within the colored books per participant is presented in Table 1. The two groups did not differ in terms of the number of words read, $t(20)=1.09$, $p=0.289$, or characters read, $t(20)=0.98$, $p=0.340$.

The behavioral Stroop results were analyzed with separate three-way ANOVAs for the reaction time (RT) and accuracy variables. Data are presented in Supplementary Table 1. Both ANOVAs consisted of the within-subjects factors of *congruency* with three levels (congruent, incongruent, and neutral) and *testing session* with two levels (pre-training and post-training) and the between-subject factor of *group* (relative of a synesthete or non-relative). Greenhouse-Geisser statistics are reported for the congruency factor in the accuracy data, because sphericity was violated. The main effect of testing session was significant for RT, $F(1,20)=6.23$, $p=0.021$, $\eta_p^2=0.239$, and for accuracy, $F(1,20)=6.35$, $p=0.020$, $\eta_p^2=0.241$. The main effect of congruency was significant for RT, $F(2,20)=6.77$, $p=0.003$, $\eta_p^2=0.253$, and for accuracy, $F(1.5,29.9)=3.80$, $p=0.045$, $\eta_p^2=0.160$. No other significant effects were found. The interaction between testing session and congruency was not significant for RT, $F(2,40)=1.92$, $p=0.160$, or for accuracy, $F(1.9,38.8)=0.06$, $p=0.946$. There was no significant interaction between testing session and group for RT, $F(2,40)<0.00$, $p=0.994$, or for accuracy, $F(1,20)=3.54$, $p=0.075$. There was no significant interaction between congruency and group for RT, $F(2,40)=0.80$, $p=0.455$, or for accuracy, $F(1.5,29.9)=0.92$, $p=0.385$. There was no significant 3-way interaction between testing session, congruency and group for RT, $F(2,40)=0.21$, $p=0.808$, or for accuracy, $F(1.9,38.8)=0.76$, $p=0.471$. Finally, there was no significant between-subjects effect of group for RT, $F(1,20)=0.05$, $p=0.835$, or for accuracy, $F(1,20)=0.29$, $p=0.595$.

The Stroop effect is defined as the difference between congruent and incongruent trial conditions (MacLeod, 1991). Based on our previous study (Colizoli et al., 2012), we expected to find a significant interaction between testing session (2 levels: pre-training vs. post-training) and Stroop congruency (2 levels: congruent vs. incongruent) in a two-way repeated measures ANOVA for the RT data. As expected, there was a significant interaction between testing session and congruency in RT, $F(1,21)=4.75$, $p=0.041$, $\eta_p^2=0.184$, but not for accuracy, $F(1,21)=0.11$, $p=0.743$. Crucially, participants responded to incongruent trials ($M=630.43$

ms, $SD=75.79$) slower than congruent trials ($M=604.59$, $SD=72.11$) after training, $t(21)=4.68$, $p<0.001$, but there was no difference between incongruent ($M=647.08$, $SD=72.20$) and congruent ($M=640.22$, $SD=81.62$) trials before training, $t(21)=0.98$, $p=0.336$. The main effect of testing session was significant for RT, $F(1,21)=7.49$, $p=0.012$, $\eta_p^2=0.263$, but was not for accuracy, $F(1,21)=3.73$, $p=0.067$. The main effect of congruency was significant for RT, $F(1,21)=12.97$, $p=0.002$, $\eta_p^2=0.382$, and for accuracy, $F(1,21)=4.80$, $p=0.040$, $\eta_p^2=0.186$. We note that including the exploratory condition consisting of the neutral trials (i.e. when including three levels of *congruency*) in the ANOVA interfered with the significant interaction between *testing session* and *congruency* in RT. Therefore, we cannot infer that behavior on the three levels, congruent, incongruent and neutral, were all widely deviating in their response to training. Instead, after training the neutral condition had response times that on average were in between the congruent and incongruent conditions (Supplementary Table 1), and the relationship in RT between congruent and incongruent trials differed significantly between testing sessions.

The Stroop effect in the RT data was also obtained in the MRI scanner after training. Participants were significantly slower to respond to incongruent trials ($M=675.59$, $SD=103.93$) compared to congruent trials ($M=631.45$, $SD=97.10$), $t(20)=2.67$, $p=0.015$. No significant difference between congruent ($M=95.60$, $SD=7.99$) and incongruent ($M=96.00$, $SD=7.99$) trials was found for accuracy, $t(20)=0.62$, $p=0.540$. Congruent trials were faster compared to the neutral trials ($M=658.11$, $SD=93.80$), $t(20)=2.08$, $p=0.051$, but were not more accurate than the neutral trials ($M=95.64$, $SD=7.71$), $t(20)=0.07$, $p=0.949$. There was no difference between the incongruent trials and the neutral trials for RT, $t(20)=1.25$, $p=0.227$, or accuracy, $t(20)=0.61$, $p=0.548$.

Despite instructions to ignore the trained color, automatic interference effects were measured by the Stroop effect after several weeks of reading books with colored letters; the Stroop effect is defined here as the measure of reaction times on incongruent trials compared to congruent trials during the post-training behavioral testing session. The Stroop effects for all participants are presented in Table 1 (range$=[-34.53, 63.78]$).

3.1.3 Vividness of visual mental imagery questionnaire (VVIQ)

Vividness of visual mental imagery was assessed using the VVIQ score. The VVIQ scores for all participants are presented in Table 1. The scores ranged from 1.34 to 2.94. The VVIQ scores were not significantly different for the two groups, $t(17.7)=0.66$, $p=0.518$ (equal variances not assumed). It should be noted that a score of 1 implies that a visual mental image is as clear and vivid as normal vision, while a score of 5 implies that there is no visual mental imagery at all. Therefore, the participants all indicated having relatively vivid visual mental imagery, because all scores were closer to 1 than to 5.

3.1.4 Projector-associator (PA) scores

After reading the colored books, participants completed a PA questionnaire designed for grapheme-color synesthetes as an exploratory experiment. Although we did not expect that the participants would necessarily report having strong associator-type or projector-type of experiences, we felt that such a questionnaire could provide a more sensitive indicator of individual differences in phenomenological experience after reading in color compared to our original questions about subjective experience (Colizoli et al., 2012, 2014a). A negative score indicates an associator-type of experience and a positive score indicates a projector-type of experience (maximum range$=[-4, 4]$). In previous studies using this questionnaire, the PA scores of synesthetes ranged from -3.1 to 4 ($n=18$) (Rouw & Scholte, 2007) and from -4 to 4 ($n=42$)

([Rouw and Scholte, 2010](#)); we note that there were partially overlapping participants in these two studies.

The PA scores for all participants are presented in Table 1. For additional interpretation of the scores of non-synesthetes on this synesthesia questionnaire, the associator and projector questions have been separately evaluated in Table 1. In the current study, the PA scores of non-synesthetes after reading in color ranged from −1.67 to 0.17. The PA scores were not significantly different for the two groups, $t(20) = 1.02$, $p = 0.320$. Ten of the 22 participants had a PA score of exactly zero. A PA score of zero could arise because the individual either (maximally) agreed or disagreed with all of the questions. Each of these ten individuals disagreed with all of the PA questions (marked a score of 1 for all questions), which means that they maximally disagreed with all associator-type and projector-type of experiences. None of the participants maximally agreed with all of the questions (marked a score of 5 for all questions), which would have represented agreeing with having both strong associator-type and projector-type of experiences. One participant had a positive score (0.17) in the projector direction, albeit very close to zero. The PA scores for the rest of the participants ranged from −1.67 to −0.17, all in the associator direction. Given the range of scores found, we can consider the PA dimension in the case of the trainees to be reflecting the degree to which the reported experiences were in the associator direction and not the projector direction. The range of scores found furthermore did not span the entire range of the associator direction. These results are notably different from the pattern of scores found for developmental synesthetes using the same questionnaire. Therefore, it is important to carefully examine the scores to arrive at a correct interpretation of the current results.

As expected, the reported experiences of acquired letter-color associations were overall related to the associator-type of experiences and not the projector-type. However, half of the sample maximally disagreed with all of the PA questions (indicated by a score of zero). Since the PA questionnaire was designed for synesthetes (who would be expected to have color experiences), a PA score of zero in this sample is ambiguous. We are not able to interpret from the PA score alone whether individuals who disagreed with the questions did so because they did not have *localizable* color experiences or because they did not have *any* color experiences evoked by the letters. In Table 1, we present the participants' responses to question 30 from the 'reading experience questionnaire' that asks participants to choose one statement out of a list of statements that best describes their experiences ([Colizoli et al., 2014a](#)). Table 1 is ordered by PA score along side the responses to this question. Overall, participants responded either that they had no color experiences related to the letters ('e' responses), or that they were internally generated experiences ('c' and 'd' responses). None of the participants chose statements related to color experiences induced by seeing the letters ('a' and 'b' responses). Of the participants with an associator score, 9 out of 11 individuals reported seeing or experiencing color in the mind's eye when they think about these letters ('c' and 'd' responses). The ten participants with a PA score of 0 all indicated they have no color experiences ('e'). Based on the converging results, we can assume that a PA score of zero in this case implies that the individual did not have any color experiences, whether they are projected into the world, in the mind's eye or a strong feeling of awareness without visual representation.

Some limitations of the comparison between the PA and reading experience questionnaires are apparent from the current data. First, there were individual differences in the semantic mapping of the verbs 'see' and 'experience' to subjective experience. All participants who chose option 'd' also agreed with projector questions, while the converse is not true. For instance, there was one trainee who had a positive PA score (0.17) who also

indicated 'no color experience.' She indicated a 2/5 Likert rating for the question '*The color is projected on the letter/number.*' Notably, this particular question lacks the terms 'see' and 'experience.' She indicated a 1/5 for all the other questions on the PA questionnaire. Due to the limited range of her responses, we would not consider her a 'projector.' Second, the reading experience questionnaire missed a category related to 'knowing' that a color is associated to a certain letter without necessarily having a visual experience (noted in [Colizoli et al. (2014a)](#)). Specifically, three of the participants who had an associator score also indicated option 'e' as a best fit for their experiences. In order to address this, we investigated their individual answers on both questionnaires. These three participants all agreed (to a degree) with the following associator questions: '*The figure does not have a color, but I am aware that it is associated with a specific color*' and '*I don't see the letters/ numbers literally in a color, but I have a strong feeling that I know which color belongs to a specific letter/number.*' None of these same participants agreed with any of the projector questions of the PA scale. It seems to be the case that these three participants probably chose option 'e' on the reading experience questionnaire because they interpreted 'see' and 'experience' to mean a visual representation of the color, while they did report 'knowing' and having a strong 'feeling' about the letter-color associations. On the reading experience questionnaire, there was no separate corresponding category for the participants who wanted to indicate 'knowing' or 'being aware' of a color association. Unfortunately, this issue was unapparent to us at the time of data collection. For this reason, the PA questionnaire was able to pick up more variance related to the subjective experience of the participants, because it described a wider range of the types of possible color 'experiences' taken in the broad sense.

3.1.5 What is the relationship between the Stroop effect, VVIQ and PA scores?

We hypothesized that individuals who report experiencing more vivid visual mental imagery would be more likely to acquire stronger letter-color associations, as measured by the size of the
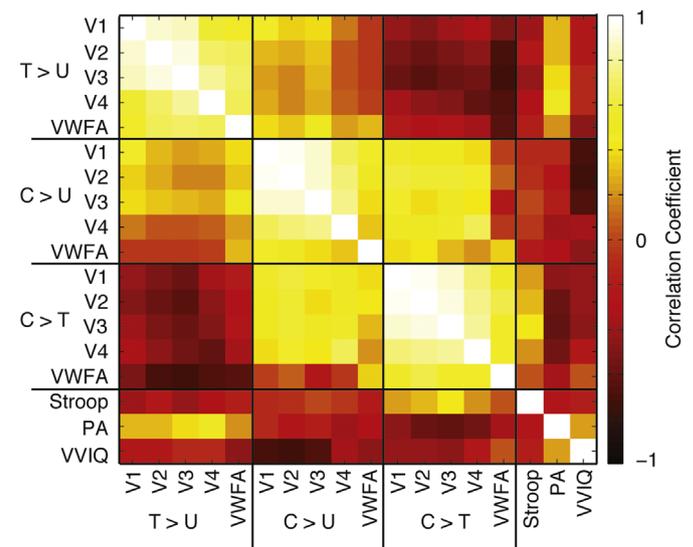
Fig. 1. The correlation matrix for variables of interest. Variables of interest were the Stroop effect (Stroop), projector-associator score (PA) and vividness of visual mental imagery score (VVIQ). Neuroimaging contrasts of interest for the color localizer were: *trained > untrained* (all achromatic) letters (T > U), *colored > untrained* letters (C > U), and *colored > trained* letters (C > T). The five brain regions of interest were retinotopically defined visual areas V1, V2, V3, V4, and the visual word form area (VWFA). All correlations reported are non-parametric rank-based correlations and were corrected for multiple comparisons using the false discovery rate.

Stroop effect after training. Correlation coefficients are presented in Fig. 1. The VVIQ score and Stroop effect variables were however not significantly correlated ($p = 0.485$), in contrast to our expectations. Additionally, individuals who reported experiencing more vivid visual mental imagery may be more likely to report having more associator-type of color experiences, as measured by the PA questionnaire after training. The VVIQ score was not significantly correlated with the PA score ($p = 0.321$). Finally, we tested whether individuals who had bigger Stroop effect sizes reported having more associator-type of color experiences, as measured by the PA questionnaire after training. The size of the Stroop effect after training was not significantly correlated with PA score ($p = 0.211$).

We therefore cannot confirm our hypotheses that the vividness of visual mental imagery or the quality of the color experience after training are directly related to the strength of the acquired associations as measured by the degree of interference of the learned letter-color association while naming incongruently colored letters. The degree to which the participants indicated localizing any color experience (mainly in the associator-range) was furthermore unrelated to the reported vividness of visual mental imagery. Taken together, these three measures seem to be reflecting different underlying factors, and therefore would be expected to differentially explain variance found in brain activation.

3.2. Neuroimaging results

The three conditions of interest during the color localizer were trained letters presented in black font color (*trained*), untrained letters presented in black font color (*untrained*), and untrained letters presented in eight different font colors (*colored*). Whenever we refer to the colored letter condition of the color localizer, we are always referring to untrained colored letters. The sets of letters used for the *untrained* and *colored* conditions were different. Contrasts of interest were *colored* > *trained* letters, *colored* > *untrained* letters, *trained* > *untrained* letters, and *untrained* > *trained* letters.

3.2.1 Whole-brain analysis of the color localizer

The contrasts of interest for the color localizer were evaluated at the level of the whole brain. None of the contrasts of interest interacted with group (*F*-tests) at the level tested. The results of the whole brain analyses for the contrasts of interest across the entire sample ($N = 22$) are reported in Table 2.

As expected, both color contrasts evoked activity along the ventral visual cortex in both hemispheres. The *colored* and achromatic letter conditions were not matched for luminance, and therefore, any activation reported includes effects related to luminance differences. The peak voxel for the contrast *colored* > *untrained* black letters was located in the right occipital fusiform gyrus and very close to the region of retinotopically defined area V4 in our sample (Fig. 2A). The peak voxel for the contrast *colored* > *trained* black letters was also located in the occipital fusiform gyrus but more anterior to both area V4 and the peak voxel of the *colored* > *untrained* black letters contrast. The contrast *trained* > *untrained* black letters did not evoke significant activation in the whole brain, while the opposite contrast *untrained* > *trained* black letters was significant in the left and right occipital lobe (Fig. 2B). For the contrast *untrained* > *trained*, the peak voxel in the right hemisphere cluster was on the border between areas V1 and V2. The peak voxel in the left hemisphere was not located in any of the retinotopically defined regions for our sample, however, the cluster itself extended into area V4.

From the contrast of letter conditions, we cannot definitively conclude that the differences in brain activation found between the *trained* and *untrained* conditions was due to training alone,

Table 2

Whole brain contrasts for the color localizer. Atlas-based brain regions, *Z*-statistic values (*Z*-max) and MNI coordinates (*x*, *y*, *z*) for the peak voxel of each cluster are reported in addition to cluster size as the number of voxels for each contrast of interest. The reported results refer to the entire sample ($N = 22$).

Cluster	Brain region	x	y	z	Z-max	Voxels
colored > untrained						
1	R occipital fusiform gyrus	32	−80	−18	5.45	2969
2	L temporal occipital fusiform gyrus	−32	−60	−18	5.35	2745
3	L posterior cingulate gyrus	−6	−26	24	3.83	83
4	L frontal pole	50	40	6	3.61	58
colored > trained						
1	R occipital fusiform gyrus	32	−74	−18	5.61	9441
trained > untrained: no significant clusters						
untrained > trained						
1	L occipital pole	−24	−96	12	4.59	674
2	R occipital pole	18	−96	8	3.94	115

because the letter conditions used each contained a different set of four letters. The letter conditions differed in respect to their exact visuospatial properties, as well as their average letter frequencies (the *trained* letters are more frequent). It is known that the frequency of letter strings differentially activates the ventral visual system (Vinckier et al., 2007). We tested if the relative activation difference found between the *trained* and *untrained* letter conditions was directly related to the amount of training. Relative BOLD activity was compared to the number of characters read in the colored books. The correlation was not significant, $r(20) = -0.13$, $p = 0.553$. Therefore, overall differences in brain activation while viewing the trained compared to the untrained black letters is likely due to effects related to both training and inherent properties of the letters. Differences in features such as low-level visuospatial properties of the letters or letter frequency may additionally contribute to BOLD-related effects of training. While a contrast with sub-optimally balanced stimuli may fall victim to such confounds, significant correlations would be robust against them, as the bias is identical for each case tested.

3.2.2 Region-of-interest analysis of the color localizer

Visual brain activation during the color localizer in response to viewing the different letter conditions across the five ROIs was investigated with a three-way ANOVA on the dependent variable of averaged normalized brain activation (*Z*-scores). The three-way ANOVA consisted of the within-subjects factors of *letter condition* with three levels (*trained*, *untrained*, and *colored*) and *visual area* with five levels (areas V1, V2, V3, V4, and the VWFA) and the between-subject factor of group (relative of a synesthete or non-relative). Data are given in Supplementary Table 2. Greenhouse-Geisser statistics are reported for the visual-area factor and the interaction between the visual-area and letter-condition factors, because sphericity was violated.

The main effect of letter condition was significant, $F(2,40) = 20.92$, $p < 0.001$, $\eta_p^2 = .511$. The main effect of visual area was significant, $F(1.9,37.4) = 37.77$, $p < .001$, $\eta_p^2 = 0.654$. The interaction between letter condition and visual area was significant, $F(3.8,75.6) = 9.28$, $p < 0.001$, $\eta_p^2 = 0.317$. No other effects were significant. Group did not interact with the within-subjects factors of letter condition, $F(2,40) = 1.09$, $p = 0.345$, or visual area, $F(1.9,37.4) = 0.47$, $p = 0.618$. There was no significant three-way interaction with group, letter condition and visual area, $F(3.8,75.6) = 0.15$, $p = 0.957$. Finally, the between-subjects factor of group was not significant, $F(1,20) = 0.68$, $p = 0.418$.
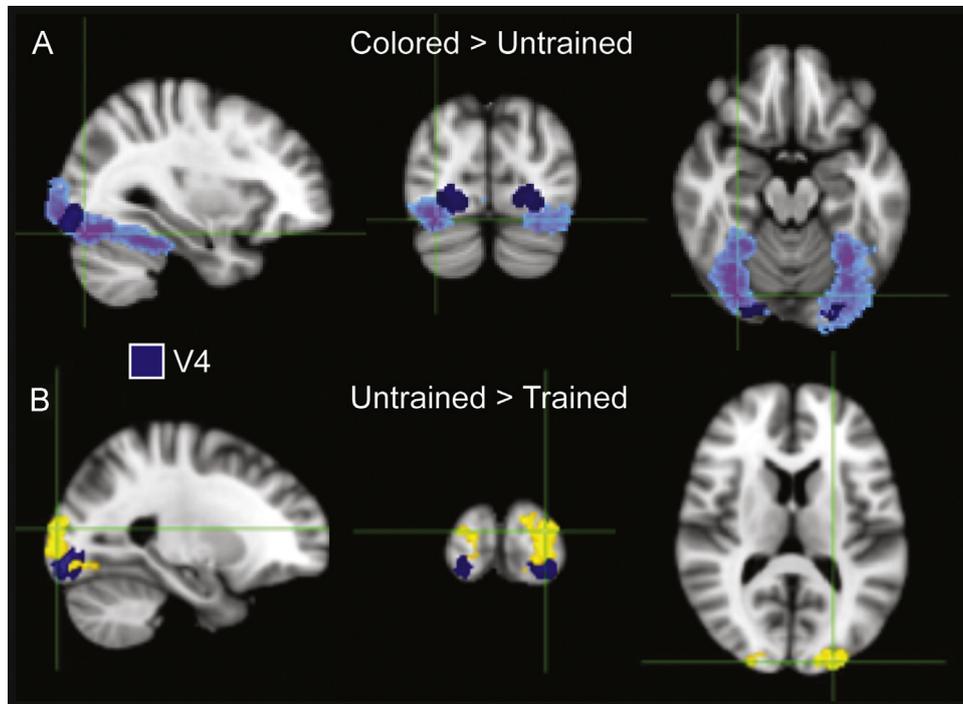
Fig. 2. Whole brain results for the color localizer and retinotopically defined visual area V4 (in dark blue). (A) Activation for the contrast *colored > untrained* letters ($32x$, $-80y$, $-18z$). (B) Average activation for the contrast *untrained > trained* letters ($-24x$, $-96y$, $12z$). The crosshairs are centered on the peak voxel of the first cluster for each contrast shown (in MNI space).

Post hoc comparisons revealed that activation while viewing the *untrained* letters ($M=0.18$, $SD=0.67$) was significantly greater compared to the *trained* letters ($M=-0.21$, $SD=0.73$), $t(21)=2.56$, $p=0.017$. Activation while viewing the *colored* letters ($M=0.74$, $SD=0.77$) was significantly greater compared to the *trained* letters, $t(21)=5.92$, $p<0.001$. Activation while viewing the *colored* letters was significantly greater compared to the *untrained* letters, $t(21)=4.26$, $p<0.001$. The difference between the *colored* and *trained* letters ($M=0.95$, $SD=0.75$) was significantly greater than the difference between the *colored* and *untrained* letters, ($M=0.56$, $SD=0.62$), $t(21)=2.59$, $p=0.017$. The difference between the *trained* and *untrained* letters was significant in each visual area except for within the VWFA (Supplementary Table 3). The difference between the *colored* and *trained* letters was significant in each visual area (Supplementary Table 3). The difference between the *colored* and *untrained* letters was significant in each visual area (Supplementary Table 3). As expected, V4 responded significantly more to *colored* than to achromatic letters (*trained* and *untrained*) compared to all of the other visual areas (Supplementary Table 3); we can thereby confirm that the paradigm was sufficiently sensitive to veridical color. The size of the difference between the *trained* versus *untrained* black letters was significantly different between areas V2 ($M=-0.31$, $SD=0.60$) and V3 ($M=-0.42$, $SD=0.54$), $t(21)=2.79$, $p=0.011$, between areas V3 and the VWFA ($M=-0.11$, $SD=0.72$), $t(21)=2.93$, $p=0.008$, and between area V4 ($M=-0.36$, $SD=0.61$) and the VWFA, $t(21)=2.53$, $p=0.035$.

As can be seen in Fig. 3, the achromatic letter conditions (*trained* and *untrained*) evoked a negative BOLD response compared to baseline (fixation cross) during the color localizer. The *trained* letters evoked a significant negative BOLD response compared to baseline in areas V1, V2 and V3 (Supplementary Table 3). The *untrained* letters evoked a significant negative BOLD response compared to baseline in area V1 only (Supplementary Table 3). Activation for the *colored* letters was not significantly different from baseline in areas V1-V3 (Supplementary Table 3). Notably, the trained letters evoked a greater negative BOLD response in areas
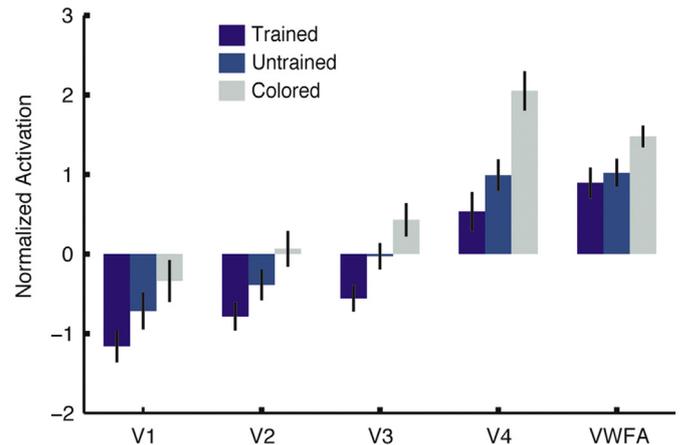
Fig. 3. Average normalized brain responses during the color localizer. The *trained* condition consisted of achromatic letters that had been consistently paired with colors during training. The *untrained* condition consisted of achromatic letters that were never paired with colors during training. The *colored* condition consisted of a different set of letters that were never paired with colors during training, presented in eight different colors (and not the same colors that had been trained). Regions of interest were retinotopically defined visual areas V1, V2, V3, V4 and the visual word form area (VWFA). Activation shown here is relative to the baseline period (fixation cross). There was a significant difference between *trained* and *untrained* letter conditions in each visual area except for the VWFA. Error bars represent the standard error of the mean of the whole sample ($N=22$).

V1, V2 and V3 compared to the untrained black letters.

3.3. Correlations between brain activation, behavior and experience

For the first time, we investigate in non-synesthetes whether brain activation in the visual cortex is related to letter-color associations acquired by reading colored books.

3.3.1 Does brain activation in visual cortex predict the Stroop effect?

Do individuals with the strongest acquired associations also show more relative visual brain activation while viewing the achromatic and colored letters? Based on the results found related to grapheme-color synesthesia by Hupé et al. (2012), we examined whether a similar difference in activation might be obtained in a training paradigm. If this is true, viewing the *trained* compared to *untrained* letters would be negatively correlated with the post-reading Stroop effect in the early visual areas V1, V2, and V3. Hupé et al. (2012) used a measure of synesthetic strength based on a combined score of two modified Stroop tasks.

We tested whether average activation in each ROI for the contrast of *trained* > *untrained* letters was related to the Stroop effect. None of the correlations were significant after correction for multiple comparisons (1-tailed); V1: $r_s(20) = -0.39$, $p = 0.035$, V2: $r_s(20) = -0.25$, $p = 0.132$, V3: $r_s(20) = -0.42$, $p = 0.027$, V4: $r_s(20) = -0.31$, $p = 0.081$, VWFA: $r_s(20) = -0.17$, $p = 0.222$. Although not significant, the directions of all correlation coefficients were negative, suggesting that a larger Stroop effect would be related to greater deactivation on average of *trained* as compared with *untrained* letters in areas V1–V3.

As an exploratory analysis, we tested whether brain activation for the two color contrasts was significantly correlated to the Stroop effect in each of the ROIs. Correlation coefficients are presented in Fig. 1. No significant correlations were found between the Stroop effect and brain activation related to the contrasts of *colored* > *trained* letters or *colored* > *untrained* letters in any of the ROIs (all $ps > 0.06$).

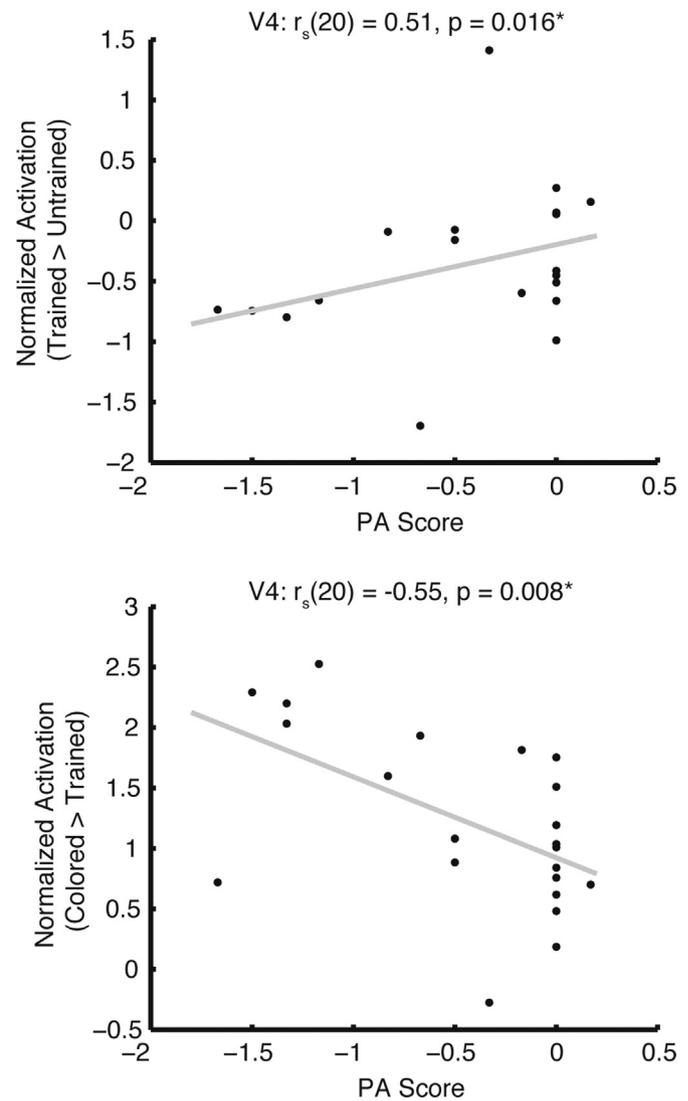### 3.3.2 Does brain activation in visual cortex predict the PA score?

Do individuals who report having stronger associator-type of color experiences also show more relative visual brain activation while viewing the achromatic and colored letters? As an exploratory analysis, we tested whether the nature of the subjective experience of color (in the absence of veridical color) as measured by PA score was related to brain activation for the three contrasts of interest within the five ROIs. Correlation coefficients are presented in Fig. 1.

PA score was significantly correlated with activation for the contrast of *trained* > *untrained* letters in area V4 only, $r_s(20) = 0.51$, $p = 0.016$ (Fig. 4A). The direction of this relationship was that individuals who reported having stronger associator-type color experiences also showed a bigger difference (in the negative direction) between the *trained* compared to the *untrained* letters in V4 (both the *trained* and *untrained* letter conditions evoked a positive BOLD response on average compared to baseline in V4 [Supplementary Table 3]). Removing the 10 individuals who had a PA score of zero (i.e. including only non-zero PA scores) increased the correlation coefficient in V4, but did not change the significance level, $r_s(10) = 0.68$, $p = 0.016$. Thus, the range of the non-zero PA scores (i.e. the individuals who reported having color experiences) was driving the correlation in V4.

Related to veridical color activation, PA score was significantly correlated with the average activation for the contrast *colored* > *trained* letters in areas V2, $r_s(20) = -0.57$, $p = 0.006$, V3, $r_s(20) = -0.62$, $p = 0.002$, and V4, $r_s(20) = -0.55$, $p = 0.008$ (Fig. 4B; only V4 is illustrated). The direction of this relationship was that individuals who reported having stronger associator-type of color experiences also showed a bigger difference (in the positive direction) between the *colored* letters compared to the *trained* letters in areas V2, V3 and V4. In contrast, no significant correlations were found between PA score and activation for the contrast of *colored* > *untrained* letters in any of the ROIs (all $ps > 0.07$).

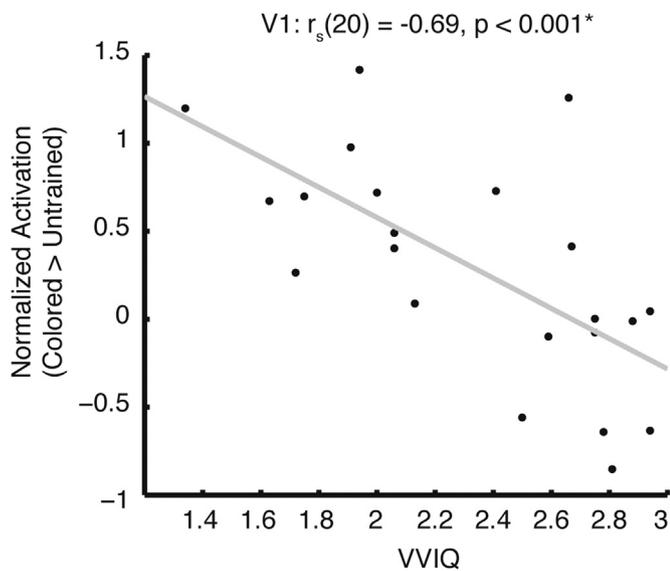### 3.3.3 Does brain activation in visual cortex predict the VVIQ score?

Do individuals who report having more vivid visual mental imagery also show more relative visual brain activation while viewing the achromatic and colored letters? As an exploratory analysis, we tested whether the reported vividness of visual mental imagery as measured by VVIQ score was related to brain activation for the three contrasts of interest within the five ROIs. Correlation coefficients are presented in Fig. 1.

No significant correlations were found between VVIQ score and activation for the contrasts of *trained* > *untrained* letters or *colored* > *trained* letters in any of the ROIs (all $ps > 0.04$). VVIQ score was significantly correlated with activation for the contrast of *colored* > *untrained* letters in areas V1, $r_s(20) = -0.69$, $p < 0.001$, V2, $r_s(20) = -0.73$, $p < 0.001$, and V3, $r_s(20) = -0.68$, $p < 0.001$. The direction indicates that individuals who reported having more vivid visual mental imagery, as indicated by a lower VVIQ score, also showed greater differences (in the positive direction) between the *colored* compared to *untrained* letters in areas V1, V2 and V3 (Fig. 5; only V1 is illustrated).

Fig. 4. The relationship between brain activation and projector-associator (PA) score in area V4. (A) PA score was significantly correlated with the average activation for the contrast of *trained* > *untrained* (all achromatic) letters in area V4 only. (B) PA score was significantly correlated with the average activation for the contrast *colored* > *trained* letters in areas V2, V3 and V4 (only V4 is illustrated here). No significant correlations were obtained for the average activation for the contrast *colored* > *untrained* letters in any of the regions of interest.

Fig. 5. The relationship between brain activation and vividness of visual mental imagery questionnaire (VVIQ) scores in area V1. VVIQ scores were significantly correlated with the average activation for the contrast *colored* > *untrained* letters in areas V1, V2 and V3 (only V1 is illustrated here). No significant correlations were obtained for the contrasts of *trained* > *untrained* letters or *colored* > *trained* letters in any of the regions of interest. A lower VVIQ score indicates having reported more vivid visual mental imagery.

4. Discussion

We have shown for the first time that reading books with consistently colored letters affects brain activation in visual cortex while viewing achromatic trained stimuli in non-synesthetes. Visual brain activation related to the contrast of *trained* versus *untrained* achromatic letter conditions was the strongest predictor of color experience measured along the projector-associator dimension of grapheme-color synesthesia. Specifically, relative activation in area V4 while viewing the achromatic *trained* letters compared to the *untrained* letters was negative in direction, and this effect significantly correlated with the PA scores (which were only in the associator-range). This suggests that area V4 reflected individual differences in how learned color associations were experienced. Furthermore, the PA scores indicated that the associated colors were considered to resemble mental images ('in my mind') more than veridical perception. In contrast, brain activation was not a reliable indicator of the objective measure of the strength of the acquired associations (i.e. the Stroop effect). Taken together, these results imply that visual brain activation is more related to the subjective experience of the acquired associations than the semantic components measured as interference effects during a reaction-time task. Related to veridical color, the difference in activation between the *colored* and *trained* letters in areas V2, V3 and V4, were related to the PA scores, while the difference between the *colored* and *untrained* letters predicted the VVIQ score in areas V1, V2 and V3. Thus, a stronger 'associator' score on the acquired color associations, and more vivid visual mental imagery, were both related to greater activation differences (in the positive direction) between the *colored* compared to *achromatic* letters, in visual brain areas.

Relative activation in early visual cortex seems to play a significant role in accounting for the individual differences in perceptual experiences of non-synesthetes exposed to frequently occurring letter-color combinations, in addition to accounting for their reported vividness of visual mental imagery. In contrast, activation in the VWFA did not correlate with any of the measures tested, indicating that the specialized grapheme-processing region

is not a neural correlate of the measured dimensions related to the acquired letter-color associations. While the contrasts of the letter conditions of the color localizer included effects of visuospatial properties and letter frequency in addition to training-related effects, there are no theoretical or empirical reasons to assume that differences between the average letter frequencies or low-level visual properties in the different conditions would contribute to the correlations between reported color experience, visual imagery and brain activation. Any confounds present in the contrasts of the color localizer were constant across all individuals tested, and therefore, would not contribute to the results obtained based on individual differences.

The cross-activation theory of synesthesia proposed that the 'epicenter' of the synesthetic color experience is located in the visual region of the brain that is known to correspond highly to the subjective experience of veridical color, namely area V4, which is also notably close to the visual word form region (Ramachandran and Hubbard, 2001a, 2001b). The literature has shown mixed results concerning the location of specific neural markers for synesthetic color in the presence of (achromatic) inducing stimuli (see: Hubbard and Ramachandran (2005), Hupé et al. (2012) and Rouw (2011)); some studies have shown that area V4 or regions near V4 were responding significantly to synesthetic color, while other studies did not. The pattern of mixed results in the literature on synesthetic color localization could arise due to several different factors that have been previously described in detail, for example, task instructions and presentation modality (Hubbard et al., 2005; Hubbard and Ramachandran, 2005; Hupé et al., 2012; Rouw et al., 2011).

In a previous study with synesthetes, projector-type of experiences were found to be related to the anatomy of sensory brain areas (Rouw and Scholte, 2010). Following this, it might be hypothesized that trained projector-like experiences are also related to stronger (training-related) activation in area V4 compared to trained associator-like experiences. In the current study, we could not test this because participants were almost entirely in the associator range of the PA dimension. The V4 activation related to associator experiences may signal that results with trainees are in contrast with the previous results with synesthetes (we thank an anonymous reviewer for pointing this out). As far as we know, this is the first study examining the neurological basis of experiences evoked by a synesthesia-like training paradigm, which means that replications are in order before strong conclusions can be drawn. Particularly because specific choices in neuroimaging analyses steer which conclusions can (and which cannot) be drawn. The current study uses only ROI-based analyses, which yields a higher signal-to-noise ratio, but is more limited in terms of brain areas investigated. In contrast, in previous synesthesia studies using whole-brain analyses, areas of visual cortex (such as area V4) were neither localized nor used as ROIs (Rouw and Scholte, 2007, 2010).

A study by Hubbard et al. (2005) was the first to employ retinotopic mapping in order to define visual area V4 in synesthetes. They found that activation for the contrast of synesthetic color was positively correlated with performance on a perceptual crowding task in areas V1, V2 and V4, indicating that individuals who performed better on this task (i.e. due to 'pop-out' of their synesthetic colors) also had greater relative activation for inducing compared to non-inducing stimuli. Such a correlation provided support for the idea that synesthetic color experiences are truly perceptual in nature, because they are directly related to activation in the visual cortex. In contrast, Hupé et al. (2012) found a pattern of negative correlations between early visual areas (V1, V2, and V3) and the strength of the synesthetic associations measured as a combined score of performance during two versions of the synesthetic Stroop task. Essentially, the measure of 'synesthesia' between both studies was qualitatively different. The synesthetic Stroop effect

(like the canonical Stroop effect) is not considered to be a measure of perceptual effects, but instead reflects a semantic component of the letter-color associations (MacLeod and Dunbar, 1988). Therefore, the inverse relationship found by Hubbard et al. compared to Hupé et al. between brain activation in retinotopically defined visual regions may reflect the different components of the synesthetic associations that were measured in each study.

Our results can be taken to suggest that activity in visual cortex is a better indication of subjective perceptual experience compared to semantic measures such as the Stroop effect. It seems logical to expect that perceptual components of the synesthetic experience would be directly related to activity in primary or secondary sensory brain regions, while the semantic or attentional components of the synesthetic associations (as measured by interferences effects of higher-level processing) would be mediated by activity in associative brain regions, such as parietal or frontal cortex. Although we did not find significant correlations between visual activity and the Stroop effect, our results for non-synesthetes were in the same direction as that found by Hupé et al. for grapheme-color synesthetes. We cannot however draw conclusions based on insignificant correlations. One point we would like to stress in order to better compare results across future studies is that a correct interpretation of negative or positive differences in BOLD activation is dependent on the relationship between the parameter estimates compared to some baseline activity in each of the ROIs. It is therefore important to show not only relative activity (e.g. *inducing graphemes > non-inducing graphemes*) but also the activity in each condition relative to the baseline used. Compared to baseline, the *trained* letters were unexpectedly related to deactivation in areas V1–V3 and activation in area V4 and the VWFA. Therefore, it was necessary to use an overall baseline condition in order to fully interpret the direction of the difference between conditions.

As expected, we found that V4 responded significantly more to the *colored* letter condition compared to any other region. If the participants were supposed to have color experiences, would we not expect the *trained* condition to look more like the *colored* condition in area V4? Activation for the contrast *colored > trained* letters was overall greater than for the contrast *colored > untrained* letters, since lower relative activation was related to the *trained* as compared to the *untrained* letter condition. As previously mentioned, these differences can be partially attributed to visuospatial and frequency differences between the letter conditions, but they cannot explain the individual differences correlations. This led us to speculate possible explanations for the underlying mechanisms. The relationship between relative activation in different brain regions may represent signals of 'competing percepts' (O'Hanlon et al., 2013). At least two studies on grapheme-color synesthesia lend some empirical support for this competing percepts hypothesis. O'Hanlon et al. (2013) found negative BOLD effects in visual cortex and in the post-central gyrus related to letters that induced synesthesia in grapheme-color synesthetes. Terhune et al. (2011) showed that inhibition of primary visual cortical activity using tDCS enhanced the synesthetic experience in grapheme-color synesthetes. Training-synesthesia paradigms are uniquely able to test such a hypothesis of competing percepts, for instance, whether certain patterns of brain activation are evident when an implicit letter-color association becomes a conscious color experience (in the absence of real color).

The self-reported vividness of visual mental imagery was predictive of the BOLD response in early visual cortex (areas V1–V3) for *colored* compared to *untrained* letters. Why would veridical color activation be related to the reported vividness of visual mental imagery? Mental imagery and veridical perception share neural mechanisms to a certain extent (McNorgan, 2012). One explanation is that individuals who have greater visual responses to salient stimuli (such as color) also have more active visual cortices in general, which are able to internally generate more vivid visual imagery. Previously, it has been shown that the VVIQ score was predictive of combined activation in areas V1 and V2 (compared to activation in the whole brain) during an imagery task (Cui et al., 2007). In that study, individuals who reported experiencing more vivid visual mental imagery also had the strongest interference effects of veridical color on a behavioral color-naming task. Similarly, Amedi et al. (2005) found a correlation between the VVIQ score and the amount of negative activation in the auditory cortex of participants during a visual imagery task. The VVIQ score therefore seems to be a reliable indicator of brain activity related to subjective reports of visual mental imagery. Interestingly, the relationship found in the current study between VVIQ score and color activation seems not to be reflected in the 'color region' of V4, but instead in earlier visual regions. Why was the relationship between color activation and VVIQ score most evident in the earlier visual areas but not in the higher areas, especially area V4? Although certain questions of the VVIQ include instructions to visualize color in the different contexts (e.g. '*the color and shape of the trees*'), these questions do not selectively address color imagery. It is possible that individuals who report experiencing vivid visual imagery have weak-to-no color imagery. Furthermore, the reported vividness of visual mental imagery was not indicative of the susceptibility to report differences in color experiences (measured as PA score) or the strength of the acquired associations (measured as the Stroop effect). Although we expected otherwise, the data suggest that the tendency to report experiencing vivid visual imagery (in general) is independent from the susceptibility to acquiring letter-color associations.

Both visual mental imagery (e.g., Barnett and Newell, 2008; Nunn et al., 2002; Rich et al., 2006) and the PA distinction (e.g., Dixon et al., 2004; Rouw and Scholte, 2010, 2007; van Leeuwen et al., 2010, 2011) have received substantial attention in the grapheme-color synesthesia literature. The PA distinction has proven to be a useful tool for capturing an essential component of synesthetic experience, because it has been able to distinguish neural markers in synesthetes (Rouw and Scholte, 2010, 2007; van Leeuwen et al., 2010, 2011). We have now shown that the PA distinction can also be relevant for characterizing certain aspects of acquired letter-color associations and furthermore is an indictor of related visual brain activation in non-synesthetes. Using the current paradigm, we could not however discern to what extent the correlations between experience and visual brain activation were related to color-specific processes or processes related to attentional selectivity for the trained stimuli. Furthermore, it was not possible to disentangle the effects of differences between the inherent visual features or letter frequency from the contrasts of the color localizer. Future research is needed in order to address these issues. In addition, it would be interesting to investigate whether the associator-type of acquired phenomenological experiences could turn into the projector-type of experiences. Perhaps it would be possible given more intense training, training over longer periods of time, or training during childhood. It is unknown how different types of training might affect brain function and structure in visual cortex. Since our participants are by definition non-synesthetes, the pattern of brain activation suggests that some common neural mechanisms may underlie letter-color associations whether they are life-long or recently acquired via training over several weeks. In the current study, we pooled together two groups of participants, the relatives of synesthetes and non-relatives, for the individual differences analyses because no group differences were obtained in any of the variables of interest in the current study. Although we did not find behavioral differences or differences in *visual* brain activation in this sample, it is possible as well as consistent with the literature that

non-synesthete relatives of synesthetes might show differences in brain function in other regions of interest, brain structure, or behavior and experience that we have not tested in the current study.

Our results illustrate the potential of how subjective experience can be affected by associative learning in addition to being reflected in brain activation. The data furthermore highlight the importance of taking into account the different dimensions of the acquired associations. Subjective experience and objective behavior such as reaction time differences are not necessarily related and are furthermore not necessarily mediated by the same brain regions entirely. Future research should aim at comparing the neurobiological markers of trainees who report having phenomenological color experiences to synesthetes in order to better assess what is uniquely related to synesthesia.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2015.07.006.

Acknowledgments

References

Amedi, A., Malach, R., Pascual-Leone, A., 2005. Negative BOLD differentiates visual imagery and perception. Neuron 48 (5), 859–872.

Asher, J.E., Aitken, M.R., Farooqi, N., Kurmani, S., Baron-Cohen, S., 2006. Diagnosing and phenotyping visual synaesthesia: a preliminary evaluation of the revised test of genuineness (TOG-R). Cortex 42 (2), 137–146.

Baker, C.I., Liu, J., Wald, L.L., Kwong, K.K., Benner, T., Kanwisher, N., 2007. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc. Natl. Acad. Sci. 104 (21), 9087–9092.

Barnett, K.J., Newell, F.N., 2008. Synaesthesia is associated with enhanced, self-rated visual imagery. Conscious. Cogn. 17 (3), 1032–1039.

Baron-Cohen, S., Wyke, M.A., Binnie, C., 1987. Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. Perception 16 (6), 761–767.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B (Methodological) 57 (1), 289–300.

Bor, D., Rothen, N., Schwartzman, D.J., Clayton, S., Seth, A.K., 2014. Adults can be trained to acquire synthetic experiences. Sci. Rep. 4, e7089.

Colizoli, O., Murre, J.M., Rouw, R., 2012. Pseudo-synesthesia through reading books with colored letters. PLoS One 7 (6), e39799.

Colizoli, O., Murre, J.M., Rouw, R., 2014a. Training synesthetic letter-color associations by reading in color. J. Vis. Exp. (84), e50893.

Colizoli, O., Murre, J., Rouw, R., 2014b. Defining (trained) grapheme-color synesthesia. Front. Hum. Neurosci. 8, 368.

Cui, X., Jeter, C.B., Yang, D., Montague, P.R., Eagleman, D.M., 2007. Vividness of mental imagery: individual variability can be measured objectively. Vis. Res. 47 (4), 474–478.

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J.B., Le Bihan, D., Cohen, L., 2004. Letter binding and invariant recognition of masked words behavioral and neuroimaging evidence. Psychol. Sci. 15 (5), 307–313.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B., Rivière, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. 4 (7), 752–758.

Dixon, M.J., Smilek, D., Merikle, P.M., 2004. Not all synaesthetes are created equal: projector versus associator synaesthetes. Cogn. Affect. Behav. Neurosci. 4 (3), 335–343.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48 (1), 63–72.

Howells, T., 1944. The experimental development of color-tone synesthesia. J. Exp. Psychol. 34 (2), 87.

Hubbard, E.M., Arman, A.C., Ramachandran, V.S., Boynton, G.M., 2005. Individual differences among grapheme-color synesthetes: brain–behavior correlations. Neuron 45 (6), 975–985.

Hubbard, E.M., Ramachandran, V.S., 2005. Neurocognitive mechanisms of synesthesia. Neuron 48 (3), 509–520.

Hupé, J.M., Bordier, C., Dojat, M., 2012. The neural bases of grapheme-color synesthesia are not localized in real color-sensitive areas. Cereb. Cortex 22 (7), 1622–1633.

Ishihara, S., 1936. The Series of Plates Designed as Tests for Colour-Blindness. Kanehara & Company, Tokyo, Japan.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17 (2), 825–841.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.

Kelly, E.L., 1934. An experimental attempt to produce artificial chromaesthesia by the technique of the conditioned response. J. Exp. Psychol. 17 (3), 315.

MacLeod, C.M., 1991. Half a century of research on the Stroop effect: an integrative review. Psychol. Bull. 109 (2), 163.

MacLeod, C.M., Dunbar, K., 1988. Training and stroop-like interference: evidence for a continuum of automaticity. J. Exp. Psychol.: Learn. Mem. Cogn. 14 (1), 126–135.

Marks, D.F., 1973. Visual imagery differences in the recall of pictures. Br. J. Psychol. 64 (1), 17–24.

McNorgan, C., 2012. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. Front. Hum. Neurosci. 6, 285.

Mylopoulos, M.I., Ro, T., 2013. Synesthesia: a colorful word with a touching sound? Front. Psychol. 4, 763.

Nunn, J.A., Gregory, L.J., Brammer, M., Williams, S.C.R., Parslow, D.M., Morgan, M.J., Gray, J.A., 2002. Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. Nat. Neurosci. 5 (4), 371–375.

O'Hanlon, E., Newell, F.N., Mitchell, K.J., 2013. Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia. Front. Psychol. 4, 755.

Polk, T.A., Farah, M.J., 2002. Functional MRI evidence for an abstract, not perceptual, word-form area. J. Exp. Psychol. Gen. 131 (1), 65.

Qiao, E., Vinckier, F., Szwed, M., Naccache, L., Valabrègue, R., Dehaene, S., Cohen, L., 2010. Unconsciously deciphering handwriting: subliminal invariance for handwritten words in the visual word form area. Neuroimage 49 (2), 1786–1799.

Ramachandran, V.S., Hubbard, E.M., 2001a. Psychophysical investigations into the neural basis of synaesthesia. Proc. R. Soc. B: Biol. Sci. 268 (1470), 979–983.

Ramachandran, V.S., Hubbard, E.M., 2001b. Synaesthesia: a window into perception, thought and language. J. Conscious. Stud. 8 (12), 3–34.

Rich, A.N., Williams, M.A., Puce, A., Syngeniotis, A., Howard, M.A., McGlone, F., Mattingley, J.B., 2006. Neural correlates of imagined and synaesthetic colours. Neuropsychologia 44 (14), 2918–2925.

Rothen, N., Meier, B., 2014. Acquiring synaesthesia: insights from training studies. Front. Hum. Neurosci. 8, 109.

Rouw, R., Scholte, H.S., 2010. Neural basis of individual differences in synesthetic experiences. J. Neurosci. 30 (18), 6205–6213.

Rouw, R., Scholte, H.S., 2007. Increased structural connectivity in grapheme-color synesthesia. Nat. Neurosci. 10 (6), 792–797.

Rouw, R., Scholte, H.S., Colizoli, O., 2011. Brain areas involved in synaesthesia: a review. J. Neuropsychol. 5 (2), 214–242.

Simner, J., 2013. Why are there different types of synesthete? Front. Psychol. 4, 558.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.

Terhune, D.B., Murray, E., Near, J., Stagg, C.J., Cowey, A., Kadosh, R.C., 2015. Phosphene perception relates to visual cortex glutamate levels and covaries with atypical visuospatial awareness. Cereb. Cortex, bhv015v1-bhv015.

Terhune, D.B., Tai, S., Cowey, A., Popescu, T., Cohen Kadosh, R., 2011. Enhanced cortical excitability in grapheme-color synesthesia and its modulation. Curr. Biol. 21 (23), 2006–2009.

van Leeuwen, T.M., den Ouden, H.E., Hagoort, P., 2011. Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. J. Neurosci. 31 (27), 9879–9884.

van Leeuwen, T.M., Petersson, K.M., Hagoort, P., 2010. Synaesthetic colour in the brain: beyond colour areas. A functional magnetic resonance imaging study of synaesthetes and matched controls. PLoS One 5 (8), e12074.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55 (1), 143–156.

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56 (2), 366–383.

Yeatman, J.D., Rauschecker, A.M., Wandell, B.A., 2013. Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. Brain Lang. 125 (2), 146–155.