Nonrandom Associations of Graphemes with Colors in Arabic

Tessa M. van Leeuwen [1,2,3,*], **Mark Dingemanse** [1], **Büşra Todil** [1,4], **Amira Agameya** [5,6] **and Asifa Majid** [1,2,7,*]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[3] Max Planck Institute for Brain Research, Frankfurt am Main, Germany
[4] Jacobs University Bremen, Germany
[5] American University of Cairo, Cairo, Egypt
[6] Cairo University, Cairo, Egypt
[7] Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

Abstract

Numerous studies demonstrate people associate colors with letters and numbers in systematic ways. But most of these studies rely on speakers of English, or closely related languages. This makes it difficult to know how generalizable these findings are, or what factors might underlie these associations. We investigated letter–color and number–color associations in Arabic speakers, who have a different writing system and unusual word structure compared to Standard Average European languages. We also aimed to identify grapheme–color synaesthetes (people who have conscious color experiences with letters and numbers). Participants associated colors with 28 basic Arabic letters and ten digits by typing color names that best fit each grapheme. We found language-specific principles determining grapheme–color associations. For example, the word formation process in Arabic was relevant for color associations. In addition, psycholinguistic variables, such as letter frequency and the intrinsic order of graphemes influenced associations. Contrary to previous studies we found no evidence for sounds playing a role in letter–color associations for Arabic, and only a very limited role for shape influencing color associations. These findings highlight the importance of linguistic and psycholinguistic features in cross-modal correspondences, and illustrate why it is important to play close attention to each language on its own terms in order to disentangle language-specific from universal effects.

---

[*] To whom correspondence should be addressed. E-mail: asifa.majid@let.ru.nl; tesvlee@gmail.com

1. Introduction

Our experience of the world is multisensory. A tomato is round, red, fragrant, sweet, and soft. An important aspect of our unconscious mental life consists of binding these distinct sensory experiences into unified perceptual experiences. We also map across senses in places where the relationships are not so obvious. For example, in many communities in Latin America there is a pervasive association of 'hot' and 'cold' with foods, drinks, natural objects, and even cultural categories (Foster, 1979). Another place we can get insight into multisensory relationships is language. For example, sounds varying in pitch are 'high' or 'low' in English, while those varying in loudness are 'soft' (*zacht*) and 'hard' (*hard*) in Dutch. Such linguistic mappings are potentially of interest, as they provide insight into how linguistic behavior can shape cross-modal associations.

In this study, we focus on associations to language in its written form. Four out of five adults worldwide are literate today (International literacy data (2014)), and writing is one of humanity's most momentous inventions (Campbell, 1997; Coulmas, 2013). Considerable research has been devoted to the cross-modal associations people have between graphemes (i.e., letters and digits) and colors, and multiple accounts of the origin and function of these associations exist. According to one account intrinsic associations exist between shapes and color (Spector and Maurer, 2008, 2011). An alternative perspective is that the linguistic environment, such as knowledge of color terms, frequency of graphemes within a language (Rich *et al.*, 2005; Simner *et al.*, 2005), or the phonetic properties of sounds (e.g., Marks, 1975) determine associations.

To date, most studies of grapheme–color associations have been in populations speaking English, using the Roman alphabet. This paper is a first investigation of grapheme–color associations for the Arabic script among Arabic speakers in Egypt. Testing grapheme–color associations in a distinct script and an independent population allows us to test the robustness of previous findings with regard to the underlying variables posited to underlie cross-modal associations (Rich *et al.*, 2005; Rouw *et al.*, 2014; Simner *et al.*, 2005).

We additionally aimed at identifying synaesthetes, a special population for whom certain stimuli are intrinsically linked to automatic cross-modal experiences (Baron-Cohen *et al.*, 1987; Hochel and Milán, 2008; Wollen and Ruggiero, 1983). Often, synaesthetic associations follow patterns of regular cross-modal associations, such as high pitch with bright colors (e.g., Ward *et al.*, 2006), but some forms of synaesthesia appear to be more arbitrary in na-

ture (Deroy and Spence, 2013). It is therefore debated whether synaesthesia can be regarded as an extreme form of normal cross-modal perception (Cohen Kadosh *et al.*, 2007; Kusnir and Thut, 2012), or not (Deroy and Spence, 2013). We explore cross-modal mappings for both synaesthetes and nonsynaesthetes in order to understand more about the mechanisms driving cross-modal associations.

2. A Closer Look at Cross-Linguistic Grapheme–Color Associations

Recently grapheme–color associations have been investigated from a cross-linguistic perspective. Simner *et al.* (2005) studied grapheme–color associations for the Roman alphabet in English and German synaesthetes and nonsynaesthetes, asking participants to freelist color associations for letters. Similar factors predicated associations across these related languages: letter–color associations were influenced by linguistic priming, i.e., letters tended to be paired with a color starting with the same initial letter (e.g., y → yellow), and in synaesthete participants, high frequency letters were associated with high frequency color words. The role of frequency has been explored in other work too (e.g., Rich *et al.*, 2005). High frequency letters and digits tend to be associated with bright colors and high saturation in synaesthetes (Beeli *et al.*, 2007; Smilek *et al.*, 2007). The correlation between digit frequency and numerical magnitude is negative in synaesthesia, that is, larger numbers are generally associated with darker colors (Cohen Kadosh *et al.*, 2007; but see Cohen Kadosh *et al.*, 2008 for an opposite effect in nonsynaesthetes).

Simner *et al.* (2005) also found lexical properties of color words affected letter–color associations. When people are asked to freelist colors in English they are likely to begin with *blue*, *red*, *green* rather than *mauve*, *olive*, or *magenta* (referred to as 'ease of generation') (Battig and Montague, 1969). Simner *et al.* found nonsynaesthetes were more likely to associate letters earlier in the alphabet with colors that were easier to generate, and those colors were also produced more often.

Another factor relevant for letter–color associations is the Berlin and Kay evolutionary sequence of colors (Berlin and Kay, 1969), which reflects the order in which colors are introduced into languages (see Malt and Majid, 2013, for a review). Although no effect was found on the order of color responses, high frequency letters were more likely to be associated with colors earlier in the evolutionary sequence in synaesthetes (Simner *et al.*, 2005).

In a similar study, Rouw *et al.* (2014) reported a strong effect of ease of generation on nonsynaesthetes' color associations for letters in English, Dutch, and Hindi, the latter employing the Devanāgarī script. The relationship between letter frequency and lexical frequency of associated color words was not tested, but overall color preferences in Hindi showed an effect of the Berlin

and Kay evolutionary sequence. Rouw *et al.* concluded speakers across the three languages and two scripts have similar color associations for similar sounds. This suggests associations may be driven by phonemes (rather than letter shape).

Evidence consistent with this comes from Japanese and Chinese, languages that combine syllabic or phonetic alphabets with a semantophonetic writing system representing meaning and sound. Chinese (L2) synaesthetes showed linguistic priming; but phonemic tone, potentially relevant for Chinese, was not found to affect color associations (Simner *et al.*, 2011). There was, however, a language-specific, or rather writing-specific effect: most Chinese characters are compounds with two components (radicals): one semantic (in the left position) and one phonetic (on the right). Hung *et al.* (2014) reported synaesthetic hue and luminance of a compound were determined by the semantic and phonetic radical, respectively.

Japanese writing combines three different types of scripts: the phonetic syllabaries Hiragana and Katakana, and the Chinese-derived Kanji. Asano and Yokosawa (2011) found synaesthetic color mapping was strongly dependent on phonology in the two phonetic scripts, Hiragana and Katakana. They proposed this was due to the strong one to one relationship between character and phonetics (unlike English where a letter can have several pronunciations; e.g., 'c' can be pronounced as /k/ or /s/).

Taken together, these studies suggest sounds of letters are especially relevant for letter–color associations. Consistent with this, English-speaking participants associate green with the letter /i/ (because of the shared sound between the letter and color word), but Arabic speaking participants did not display this association (Guillamon, 2013). However, across languages, including Arabic, front open vowels were often associated with red, and front mid vowels with green (Table 5 in Guillamon, 2013). According to Marks (1975), the pitch of vowels predicts the lightness/darkness of the associated color (e.g., 'o' and 'u' are associated with darker colors, and sound lower in pitch, compared to 'e' and 'i').

Another factor that appears to be relevant for letter–color associations is letter shape. For example, Watson *et al.* (2012) found letter shape (measured by 11 letter shape features) and ordinality (position in the alphabet) correlated with hue and color distance in synaesthetes, while letter frequency correlated with luminance. Some synaesthestic participants also associated similarly shaped graphemes with similar hues (Brang *et al.*, 2011; Jürgens and Nikolić, 2012; Watson *et al.*, 2010). Similarly, Spector and Maurer (2008, 2011) found preliterate toddlers systematically associate, e.g., 'o' with white and 'x' with black, according to letter shape.

In sum, grapheme frequency, sound, and shape have been linked to color associations, at least in work based mostly on the Roman alphabet. Here we examine which factors are relevant for grapheme–color associations in Arabic.

3. Arabic and the Arabic Script

Arabic is an Afro–Asiatic language closely related to other Semitic languages such as Hebrew and Aramaic. It is the official language of the 22 countries in the Arab World. Spoken by around 240 million people, it is the fifth most populous first learned language worldwide (Lewis *et al.*, 2014).

Spoken Arabic differs considerably depending on the region; but spoken varieties are considered mutually intelligible and form a dialect chain. Egyptian Arabic, which we study here, is probably the most widely understood spoken variety. Its written counterpart, Modern Standard Arabic, is the only official form of Arabic, and is used in most written documents. The Quran is written in Classical Arabic. Literacy in the Arab States is around 77.5%.

Arabic is written using a right to left script consisting of 28 letters (Fig. 1A), with digits written from left to right (Fig. 1B; we refer to these as Eastern Arabic numbers to prevent confusion with the 'Arabic numbers' 0–9). All Arabic letters are consonants, and short vowels can optionally be indicated using dia-

A

ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا
S	sh	s	z	r	dh	d	x	H	j	th	t	b	'aa
Saad	shiin	siin	zaay	raa'	dhaal	daal	xaa'	Haa'	jiim	thaa'	taa'	baa'	'alif

ي	و	ه	ن	م	ل	ك	ق	ف	غ	ع	ظ	ط	ض
y	w	h	n	m	l	k	q	f	gh	[ʕ]	Z	T	D
yaa'	waaw	haa'	nuun	miim	laam	kaaf	qaaf	faa'	ghayn	'ayn	Zaa'	Taa'	Daad

B

٠	١	٢	٣	٤	٥	٦	٧	٨	٩
0	1	2	3	4	5	6	7	8	9
Sifr	waaHid	ithnaan	thalaatha	'arba'a	xamsa	sitta	sab'a	thamaaniya	tis'a

Figure 1. Arabic alphabet and digits. Panel A depicts the Arabic alphabet. The letters should be read right to left. 'alif ا is the first letter of the alphabet. Corresponding sounds are given in the light grey panel, and the Arabic letter name is in the dark grey panel. Panel B depicts Arabic digits (also known as Eastern Arabic numbers). The numbers are read from left to right. The top line shows the Arabic digits used in the study. The corresponding English numbers follow in the light grey panel. The Arabic names for the numbers are given below in the dark grey panel. Adapted from Ryding (2005).
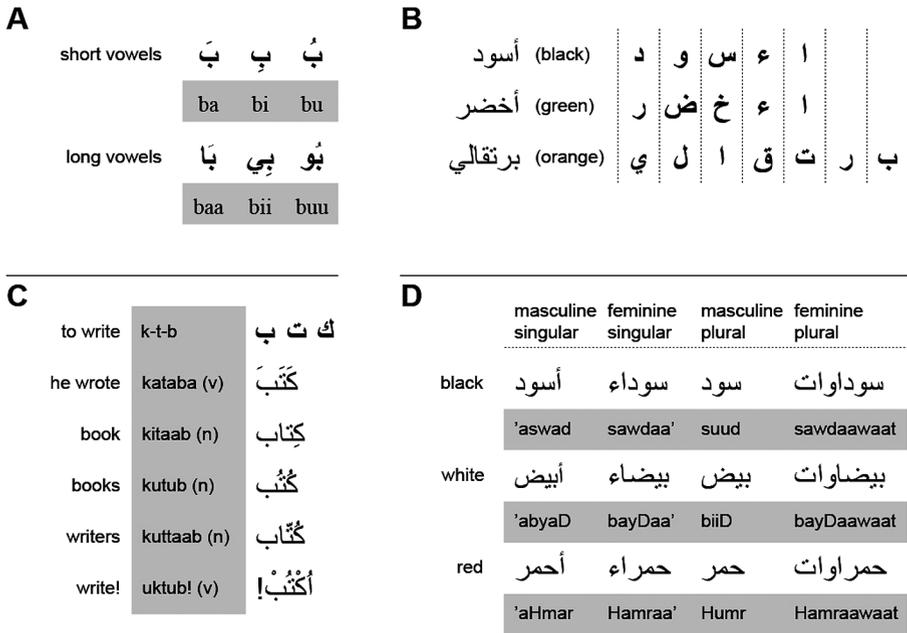
A

short vowels   بَ   بِ   بُ

ba   bi   bu

long vowels   بَا   بِي   بُو

baa   bii   buu

B

أسود (black) | د | و | س | ء | ا |

أخضر (green) | ر | ض | خ | ء | ا |

برتقالي (orange) | ي | ل | ا | ق | ت | ر | ب |

C

to write	k-t-b	ك ت ب
he wrote	kataba (v)	كَتَبَ
book	kitaab (n)	كِتاب
books	kutub (n)	كُتُب
writers	kuttaab (n)	كُتّاب
write!	uktub! (v)	أُكْتُبْ!

D

	masculine singular	feminine singular	masculine plural	feminine plural
black	أسود	سوداء	سود	سوداوات
	'aswad	sawdaa'	suud	sawdaawaat
white	أبيض	بيضاء	بيض	بيضاوات
	'abyaD	bayDaa'	biiD	bayDaawaat
red	أحمر	حمراء	حمر	حمراوات
	'aHmar	Hamraa'	Humr	Hamraawaat

Figure 2. Details of Arabic script. Panel A shows how vowels can be represented in Arabic using diacritics, or in combination with three of the letters from the alphabet. Panel B shows how Arabic letters change shape when written in words. Three color word participants produced in the study are shown on the left side, with the corresponding fully written out letters on the right side. Panel C gives an illustrative example of the rich morphology of Arabic. The Arabic word is shown on the right, with the transliteration and word class (n = noun; v = verb), and translation. Panel D shows how three of the color words change their form as a function of grammatical gender and number.

critics above or below letters (Fig. 2A). Three of the letters used for consonants (ا 'alif, ي yaa', and و waaw) are also used for long vowels. Diacritics are usually omitted in written Modern Standard Arabic (as in Fig. 2B and D).

Letters change their appearance when they are written in words, as can be seen in Fig. 2B. Twenty-two of the letters can be connected to the preceding and following letters, but six letters can only be joined to the preceding letter, not the following letter, e.g., ا 'alif and و waaw.

Arabic word structure is different to English and other related languages. The core of an Arabic word is the 'root', which usually comprises three (range 2–5) consonants and expresses the main conceptual content. For example, the root *k-t-b* (ك - ت - ب) expresses the meaning 'to write' (see Fig. 2C). These consonantal roots serve as templates with slots for different vowel patterns to form different words, as illustrated in Fig. 2C. So, the root *k-t-b*, can become *kataba* 'he wrote'; *kitaab* 'book', *kutub* 'books', kuttaab 'writers', or the com-

mand *uktub* 'write!'. There are around 5000–6500 different lexical roots in Arabic (Ryding, 2005).

This word formation process (known as templatic morphology) has potentially interesting implications for color associations. Prior studies have shown people associate Roman letters (e.g., *y*) with colors beginning with the same letter (e.g., *yellow*). In Arabic, however, the same color concept, e.g., 'black' can be expressed with word forms that begin with different letters (Fig. 2D). So, 'black' can begin with ا *'aa* or س *s* depending on the gender and number of the corresponding noun (the adjective has to agree with the noun). As well as the adjectival forms, there is also a verb سود 'to blacken'. This raises the question whether Arabic speakers associate letters with colors on the basis of the surface word form or on the basis of the underlying root.

Many Arabic letters share a strong shape resemblance (Elbeheri and Everatt, 2006), in part since several otherwise identical characters are distinguished only by the placement of a disambiguating dot (see Fig. 1A). Previous studies have found English synaesthetes to be influenced by shape similarities among letters in their color associations, i.e., similarly shaped letters are associated with similar colors (Brang *et al.*, 2011; Jürgens and Nikolić, 2012; Watson *et al.*, 2010). Given the number of shared shapes in Arabic orthography, we predict shape plays an important role in determining color associations in Arabic as well.

4. Current Study

In the current study, we assessed the influence of linguistic and other factors on grapheme–color mapping in the Arabic language. We address the influence of grapheme frequency, ease of generation order, the Berlin and Kay evolutionary sequence, matching between graphemes and the first letter onset of the chosen color word, grapheme shape and sound, as well as the order in which stimuli are presented to the participants. Following Simner *et al.* (2005), we presented stimuli in written form and asked participants to write down the names of colors they thought matched each grapheme.

4.1. Methods

4.1.1. Participants
Participants were recruited to participate in a 15 min online survey *via* university mailing lists in collaboration with the English Language Institute at the American University of Cairo, Egypt. Participants were not paid. All participants were informed of the content of the survey and formally asked to accept the conditions (informed consent). Ten demographic questions regarding participants' age, gender, origin, language information, keyboard use (Arabic or English), handedness, and color blindness were administered.

In total, 139 participants completed the study; however, the demographic questions were completed by only 121 participants (the questions were placed after the first session of the questionnaire and several people quit at this point). We can therefore only address the demographic characteristics of these 121 participants. Their mean age was 25.3 years (SD = 7.8; 25 males, 96 females, 10 of unknown gender). Native languages were Arabic and English (55%), Arabic only (25%), and Arabic in combination with one or more European languages of which at least one was not English (17%). Languages of daily use were Arabic and English (79%), Arabic only (11%), and Arabic in combination with one or more European languages of which at least one was other than English (8%). Eighty-eight participants were born in Egypt (73%), and eleven participants originated from non-Arabic speaking countries, while the remaining participants came from other Arab countries. Sixteen participants reported themselves as left-handed or ambidextrous (13%), while two people reported color blindness. On a scale from 1 to 5 of keyboard use (1 being only English/Roman keyboards and 5 being only Arabic keyboards), the participants scored 1.73 on average (SD = 0.75).

4.2. Structure of the Online Survey of Grapheme–Color Mappings

Participants were presented with the 28 letters of the Arabic alphabet (no diacritics) and ten Eastern Arabic digits (see Fig. 1) and typed the name of the color that best fit each grapheme. Keyboards had both English and Arabic characters on the keys. This method allows us to investigate the psycholinguistic factors underlying grapheme–color associations. Immediately after completing the task, participants were asked to do the same association task again. The repetition was essential to test the consistency of color associations and possible identification of synaesthetes. The stimuli were presented either in fixed alphabetic/numerical order (in both repetitions) or random order (in both repetitions), as the presentation order of stimuli can influence association results, e.g., because of ease of generation effects (see Simner *et al.*, 2005). The survey started either with digits or letters (in both repetitions). Hence there were four different versions of the survey: Fixed stimulus order starting with letters, fixed stimulus order starting with digits, random stimulus order starting with letters, and random stimulus order starting with digits. Assignment of the survey version to participants at the start of the survey was done at random as programmed in the survey software (Unipark, http://www.unipark.com/). The survey instruction language was Modern Standard Arabic.

4.3. Identification of Synaesthetes

After the color associations followed six questions with a Likert rating scale of 1 to 6 (do not agree – agree) to assess the extent to which participants' experiences resembled synaesthesia. These questions were taken from the prevalence

study of Simner *et al.* (2006). An example statement is: '*When performing the experiment, I felt that I knew for certain what the color for a letter or number should be*'. The minimum total score was 6 (not resembling synaesthesia) and the maximum score 36 (resembling synaesthesia). Across all participants, we calculated the average rating score on these six questions, and if a participant's score exceeded a cutoff value of (mean + 1.96 SD), this was considered an indication for synaesthesia. Additionally, consistency in the color associations was judged from the immediate retest of the color associations in the online survey. For the letter and number surveys, we calculated the average number of color associations that were the same between test and retest across all participants. If, for a particular participant, the number of identical color responses between test and retest ('consistency score') exceeded the cutoff value (mean + 1.96 SD), this was considered an indication for synaesthesia. Combined with the responses to the synaesthesia-related questions this allowed us to discern synaesthetes from nonsynaesthete participants.

In earlier studies aiming to identify synaesthetes, synaesthetes were taken to be those individuals falling within a range of scores previously established for synaesthetes (on both consistency and questionnaire scales) (Rothen and Meier, 2010; Simner *et al.*, 2006). Since no known synaesthesia scores are available for individuals using the Arabic script, we instead used a conservative estimate by requiring potential synaesthetes to exhibit a score above the specified cutoff values.

5. Results

In total, 139 participants completed at least one session of the questionnaire. Even though the survey was presented in Modern Standard Arabic, only 92 participants responded in Arabic; 47 responded in English. This high percentage of English responders could be related to the high percentage of participants (79%) who used English alongside Arabic in their daily life. Arabic and English responses were separated for all analyses. Of the Arabic responders ($N = 84$), 51 participants completed both repetitions of the stimulus list (61%) for English responders ($N = 39$), 19 completed both runs (49%).

Arabic responses were transliterated using AraMorph (Buckwalter, 2002), and Aralex (Boudelaa and Marslen-Wilson, 2010) was used to extract roots (e.g., برقل for the Arabic word برتقالي orange color), and obtain transliterations of the surface forms and roots (e.g., brtqAly, brql), corpus frequency of the surface Arabic forms, and root token frequency. For Arabic words not listed in AraMorph or Aralex, a gloss was obtained from Google Translate (http://translate.google.com) and checked with author AA, who is a native Arabic speaker.

After translation, but before further analyses, the data underwent pre-processing. Participants who answered only 'black' or 'no color' for each grapheme were removed from the analysis because there was no variation in their color responses (eight participants for Arabic, eight for English). Synaesthetes were identified, and analyzed separately from the other participants (see below). Arabic color responses for which no frequency information could be obtained from Aralex were not considered in the frequency analyses, but were used for identification of synaesthetes.

5.1. Identification of Synaesthetes

Synaesthetes were identified separately for Arabic and English, because language and cultural factors may influence the choice of colors and responses on questionnaires. Also, consistency scores for letter and number responses were analyzed separately in order to independently identify letter–color and number–color synaesthetes. The synaesthesia questionnaire and consistency scores and cutoffs are summarized in Fig. 3 and Table 1. Two synaesthetes were identified (both female); additionally, one participant was close to the criteria for synaesthesia for Eastern Arabic numbers. Data from the two synaesthetes were removed from further analysis. The synaesthetes' responses could not be analyzed in detail because of the small sample size (although see Section 5.7).

For Arabic responders, the synaesthesia questionnaire score and the consistency score correlated strongly and positively for letters ($r = 0.562$, $p < 0.001$) as well as for Eastern Arabic numbers ($r = 0.490$, $p < 0.001$), suggesting the two measures are related. For English responders, this relationship was not significant (letters: $r = 0.164$, $p = 0.545$; numbers: $r = 0.352$, $p = 0.198$).

5.2. Grapheme–Color Pairings Are Not Random

In order to ascertain whether grapheme–color combinations were systematically associated we first counted the total number of times a color was chosen as a response. From this, we derived the probability a given grapheme would be associated with a color by chance which served as a baseline for calculating binomial probabilities (for instance 14% of all responses consisted of the color red). We then counted the actual number of times colors were chosen for each grapheme, and used the binomial probability to calculate how likely these grapheme–color pairings were if the distribution would be random. The derived p values were Bonferroni-corrected for multiple comparisons because of the number of colors tested (see Simner *et al.* (2005) and Ward and Simner (2003)).

For Arabic and English responses to letters, color words outside of the eleven basic color terms were frequent enough to be considered as separate
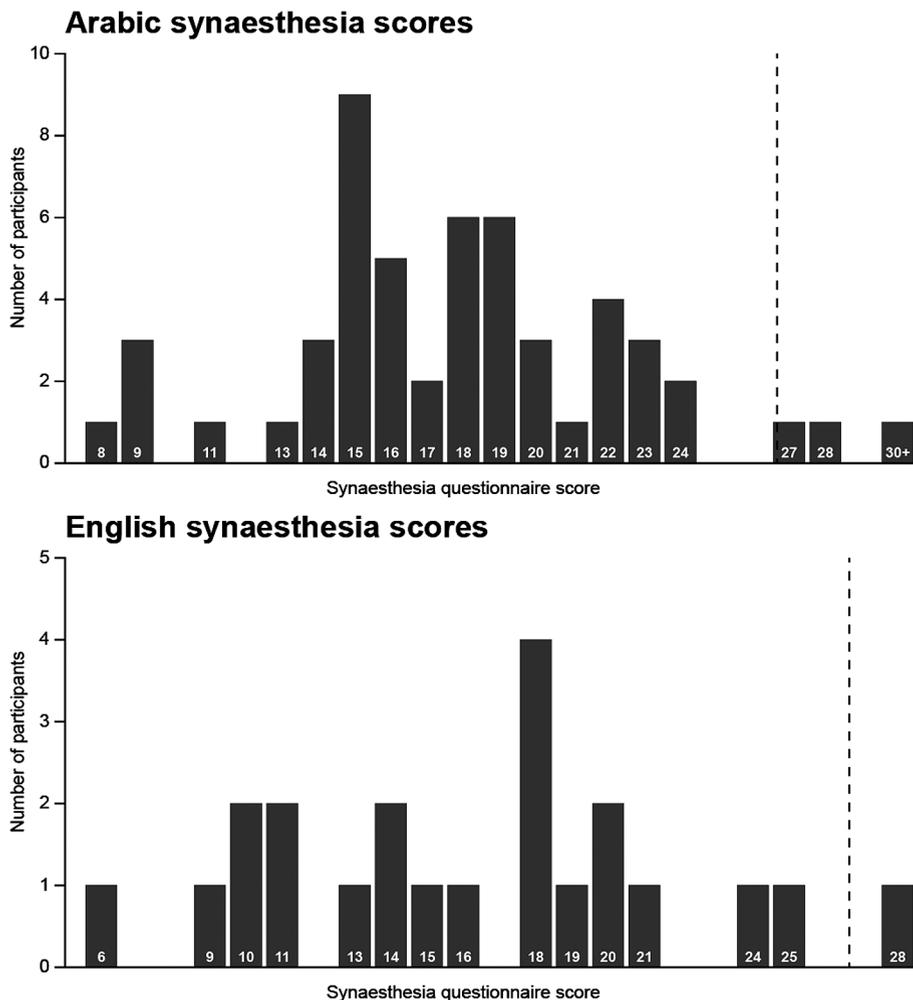
Arabic synaesthesia scores

English synaesthesia scores

Figure 3. Synaesthesia prevalence and scoring distributions. Distribution of scores on the synaesthesia questionnaire for Arabic and English respondents.

colors for analysis. For Eastern Arabic numerals, nonstandard colors made up only 3.9% of the total amount of responses and those nonstandard color responses were collapsed onto one of the eleven basic color terms (black, white, gray, brown, red, orange, yellow, green, blue, purple, pink) (Berlin and Kay, 1969). Violet was collapsed onto purple, silver onto gray, golden onto yellow. For English responses to numbers, the nonstandard responses made up 3.1% of the total and the same procedure was followed.

The significant ($p < 0.05$) grapheme–color pairings are summarized in Figs 4 and 5. Some regularities are particularly noteworthy. For example, the

Table 1.
Synaesthesia

	Synaesthesia questionnaire responses (max. score = 36)		Consistency score*		>cutoff quest	>cutoff consist	Synaesthetes
Arabic	completed:	N = 53					
	mean:	17.7 ± SD = 4.81					
	cutoff:	27.1					
	letters (N = 51)**		mean:	24.2 ± SD = 12.6	N = 2	N = 1	N = 1
			cutoff:	48.8			
	numbers (N = 47)		mean:	8.55 ± SD = 5.61	N = 0***	N = 1	–
			cutoff:	19.54			
English	completed:	N = 23					
	mean:	16.3 ± SD = 5.53					
	cutoff:	27.1					
	letters (N = 19)		mean:	17.0 ± SD = 12.8	N = 1	N = 0	–
			cutoff:	42.2			
	numbers (N = 18)		mean:	8.67 ± SD = 5.44	N = 1	N = 1	N = 1
			cutoff:	19.3			

* Letters: maximum score 56, numbers: maximum score 20.
** Number of participants that completed both sessions and for whom a consistency score could be calculated.
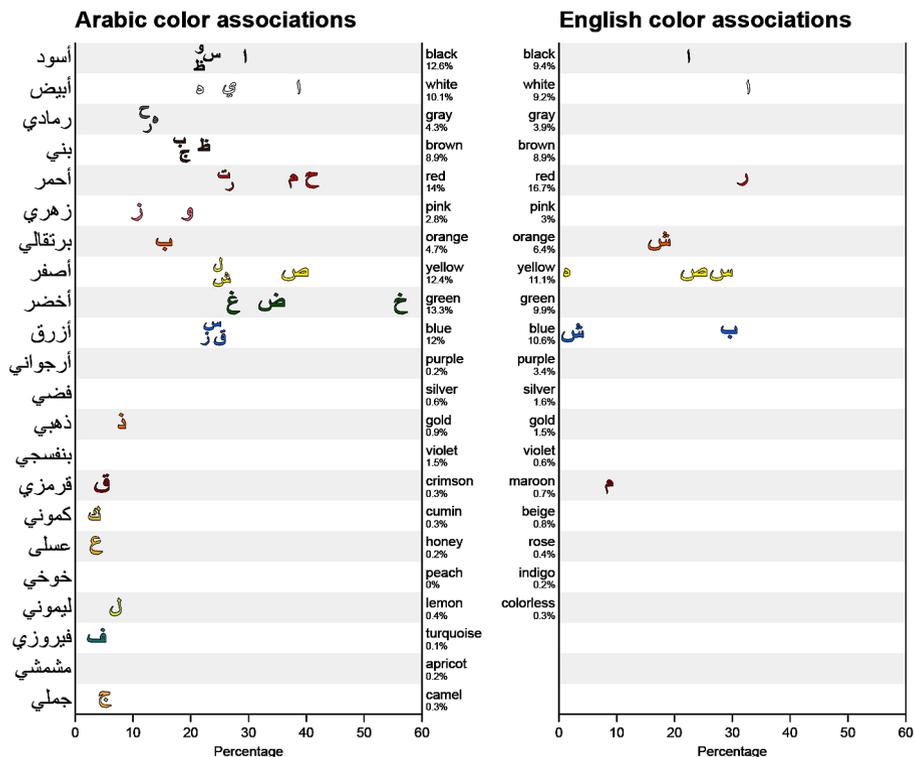*** One participant reached a score of 27, just below the cutoff of 27.1.

Figure 4. Letter–color associations for Arabic and English respondents. Left panel shows the above chance Arabic color responses to Arabic letters. English translations for the Arabic words are shown on the right hand side, with their corresponding chance values. English color associations to Arabic letters are shown in the right panel. This figure is published in color in the online version.

letter 'alif ا, the first letter in the Arabic alphabet, was associated with black and white significantly more often than chance (both colors $p < 0.001$). This was true for both Arabic and English responders. Intriguingly, a similar pattern holds for the numbers 0 ٠ and 1 ١ (0 is associated with black at $p < 0.001$ and white at $p < 0.01$; 1 is associated with black at $p < 0.01$ and white at $p < 0.001$). Earlier studies (Beeli *et al.*, 2007; Smilek *et al.*, 2007) found similar associations between 0 and 1 with black and white in English speaking synaesthetes, but not for the first letter of the alphabet. Rich *et al.* (2005) find that number 1 is significantly associated with black or white (0 was not included in their study) for nonsynaesthetes and white for synaesthetes. There are two things to note. First, it is probably not a coincidence that there is a strong similarity in shape between the letter 'alif and the number 1. Second, these associations are present in the nonsynaesthetes in our study, unlike most
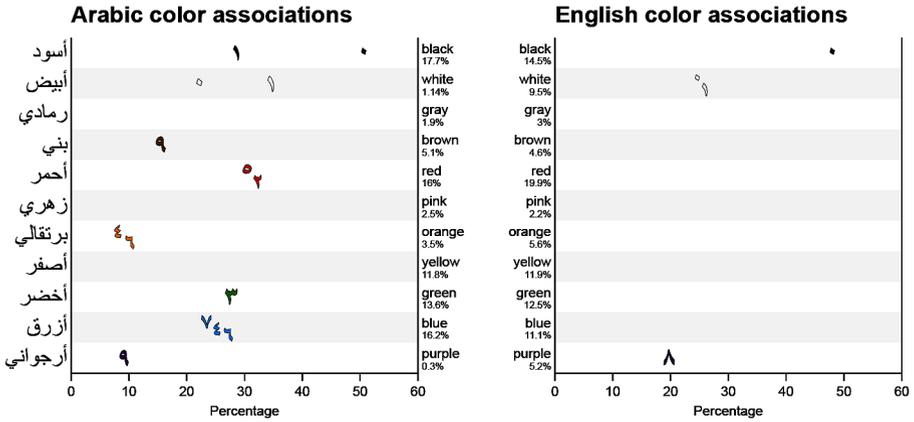
Figure 5. Digit–color associations for Arabic and English respondents. Left panel shows the above chance Arabic color responses to Eastern Arabic numbers. English translations for the Arabic words are shown on the right hand side, with their corresponding chance values. English color associations to Eastern Arabic numbers are shown in the right panel. This figure is published in color in the online version.

previous English studies which find the effect only with synaesthetes (with the exception of Rich *et al.*, 2005).

Other letter–color associations were similar across both groups of respondents. For example, ر *r* was associated with red by both Arabic and English respondents. This was also the case for ص *S* and yellow. The letter م *m* was associated with red by Arabic respondents, and with a subordinate of red, i.e., maroon, by English respondents. However, there were many differences between the groups too (see Figs 4 and 5). For example, three different letters had very strong associations with green amongst the Arabic respondents (all $p < 0.001$), but no English respondent associated any letter with green above chance levels. It is tempting to speculate this prominent association reflects the cultural significance of green for Arabic respondents. Green is the traditional color of Islam.

5.3. Arabic Letters Match Onsets of Color Words and Their Roots

We also investigated whether the onset letters of Arabic color word responses (and their roots) corresponded with particular Arabic graphemes more often than expected by chance. This analysis is only relevant for Arabic responses to letter stimuli. For each color word response in the Arabic letter condition, we noted the onset letter of the raw color word response produced, and the onset letter of the root of the response word. The onset letter of the raw color word matched the given stimulus letter in 8.6% of cases (when randomizing the data, onset matches occurred in 2.9% of cases). For the root of the response

word, the percentage of matches was 12.0% (for random data, root matches occurred in 6.7% of cases). Overall, then, this analysis suggests there was a higher than chance tendency to associate color words with both raw onsets and root onsets of the experimentally presented letters, although this was not quantified statistically.

In addition, we tested whether specific onset letters of the Arabic color word and the onset letter of the root corresponded to a given letter more often than expected by chance, using the same binomial probability method described above (e.g., did ب *baa'* elicit the response بني *bunni* 'brown'). For the raw Arabic color word responses in the letter condition, an onset match in the response occurred for the grapheme waaw (و) more often than chance ($p < 0.05$), but for no other specific grapheme–onset combination.

5.4. *Grapheme Frequency Influences Color Word Choices in Arabic and English Responders*

To test the relationship between grapheme frequency and color word frequencies, linear regression analyses were performed with Arabic letter and digit frequency as the independent variable and frequencies of the (Arabic and English) color responses as a dependent variable. Participants were included as a random factor. Arabic letter frequencies were derived from Intellaren.com (Note 1), a 1.3 million word Arabic corpus (approximately 5 million letters) and are listed in Table 2. Letter frequencies can be calculated lumping (or not) modified letter forms onto their primary forms. We only presented primary letter forms ($N = 28$); therefore we used nonlumped letter frequencies. No explicit source was found for Arabic digit frequency, but based on data from other languages the ranked digit frequency was taken to be the inverse of the digit number (e.g., Beeli *et al.*, 2007). As stated above, Arabic color word frequencies and properties were derived from Aralex and are listed in Table 3. Ranked English color word frequencies were derived from the British National Corpus, as listed in the Appendix of Simner *et al.* (2005).

When analyzing Arabic color word responses with respect to their orthographic word frequency, we noticed the word frequency for 'brown' (بني in Arabic) as derived from Aralex was uncharacteristically high compared to other color word frequencies (43.15 compared to values in the range of 0–10). The word بني is ambiguous and can be derived from the root 'to build' or 'sons of'. Other color words are also ambiguous in their written form, but did not cause this magnitude of distortion. Aralex reflects orthographic frequency as encountered by Arabic readers. Still, given the discrepancy of orthographic

Table 2.
Letter frequencies in Arabic

Arabic letter	Letter frequency*
ا	12.5
ل	12.07
ن	6.61
م	6.52
ي	6.36
و	5.8
ه	5.08
ب	4.67
ر	4.2
ع	4.01
ف	2.84
ق	2.69
د	2.67
ت	2.61
س	2.47
ك	2.01
ح	1.86
ج	1.23
ص	1.04
ذ	0.96
ث	0.87
خ	0.79
ش	0.73
ز	0.52
ط	0.5
ض	0.44
غ	0.33
ظ	0.18

* Derived from Intellaren.com.

frequency of بني in comparison to other words, and the fact the high frequency is most likely due to its additional senses, we excluded بني responses from the analyses reported next.

Table 3.
Color word frequencies in Arabic

Raw Arabic form	Color	Orthographic frequency Arabic form Aralex
مشمشي	apricot	x*
أسود	black	7.8
اسود	black	4.19
أزرق	blue	1.92
ازرق	blue	1.09
بنى	brown	43.15
قرمزى	crimson	0.23
كموني	cumin	0.08
دهبي	golden	1.85
رمادي	gray	0.83
أخضر	green	2.86
اخضر	green	1.35
عسلى	honey	x*
ليموني	lemon	0.03
برتقالى	orange	0.39
خوخي	peach	0.03
وردي	pink	1.43
أرجواني	purple	0.03
بنافسجى	purple	0.13
أحمر	red	4.97
احمر	red	2.18
فضي	silver	0.44
فيروزي	turquoise	0.26
بنفسجي	violet	0.13
أبيض	white	6.94
ابيض	white	3.88
أصفر	yellow	1.17
اصفر	yellow	0.57

* x denotes color word for which no frequency information was listed in Aralex.

We found a strong positive relationship between the frequency of occurrence of Arabic letters and orthographic frequency of associated Arabic color words ($F[1, 2975] = 22.8$, $p < 0.001$, $R^2 = 0.008$, $\beta = 0.087$, $t = 4.77$). The same positive relationship was found between the frequency of the Arabic letter and the root token frequency of the chosen color word ($F[1, 2451] = 18.8$, $p < 0.001$, $R^2 = 0.008$, $\beta = 0.087$, $t = 4.34$) (Note 2). For English responses, the color word frequency was taken to be the ranked frequency of color words in the English language. We found the same relationship: high frequency Arabic letters were associated with high frequency color words in English ($F[1, 1546] = 9.85$, $p < 0.01$, $R^2 = 0.006$, $\beta = 0.080$, $t = 3.14$).

Similar positive correlations were found for Eastern Arabic number frequency and associated Arabic color word frequency ($F[1, 1094] = 81.2$, $p < 0.001$, $R^2 = 0.069$, $\beta = 0.263$, $t = 9.01$); number frequency and root token frequency ($F[1, 871] = 38.2$, $p < 0.001$, $R^2 = 0.042$, $\beta = 0.205$, $t = 6.18$), and number frequency and rank order of the frequency of English color words ($F[1, 508] = 31.2$, $p < 0.001$, $R^2 = 0.058$, $\beta = 0.272$, $t = 5.59$).

Figure 6 illustrates this effect for responses in Arabic with a high frequency color word (white squares) and a relatively low frequency color word (violet triangles). High frequency letters are associated with the high frequency color word, whereas for lower frequency letters, lower frequency color words are produced.
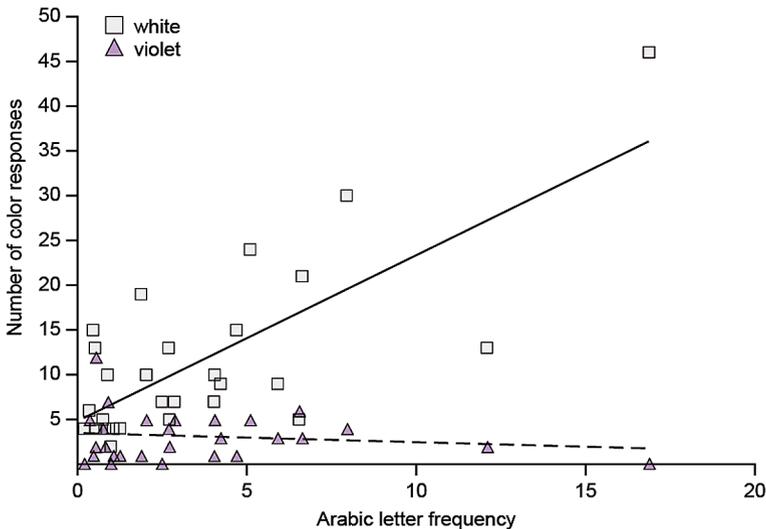

Figure 6. Relationship between letter frequency and color words associated. Higher frequency color words (e.g., white) were more likely to be associated with high frequency Arabic letters, whereas lower frequency words (e.g., violet) were more likely to be associated with lower frequency letters. This figure is published in color in the online version.

5.5. *Similarly Shaped or Sounding Letters Are Not Systematically Associated with Similar Colors*

Earlier we saw that the letter 'alif and number one received similar color associations, i.e., black and white. This suggests similarity in shape might influence color associations. Figure 7 plots letters with similar shapes together with their color associations. The results do not support the hypothesis that color associations are driven by shape similarity. Only the letter pair ط *T* and

ظ *Z* showed a potential similarity in color association, i.e., brown (although

note only the association of brown with ظ *Z* was significantly above chance).

This suggests shape has only a limited role to play in color associations, perhaps limited only to the first letter and digit of Arabic.

We also investigated whether letters that sound similar in Arabic tend to be associated with similar colors (Fig. 8). The letters were plotted according to Ryding (2005, p. 13), and the significant color associations are shown by a colored square around the letter. As with letter shape, there was no clear pattern between sound features and color associations in this task.

5.6. *Color Responses Are Influenced by the Berlin and Kay Evolutionary Sequence*

We assessed whether the order in which color word responses were generated by participants was random or not. According to the Berlin and

	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	
أسود		+									*				*						black
أبيض																					white
رمادي					*				*												gray
بني	*			*			+								+	*					brown
أحمر		*		*					*												red
زهري										*											pink
برتقالي	*																				orange
أصفر												*	*								yellow
أخضر						*								*			*				green
أزرق											*	*							*		blue
أرجواني																					purple
ذهبي						*															gold
عسلى																	*				honey
فيروزي																			*		turquoise

Figure 7. Letter shape and color associations. Letters are depicted in groups based on similarity of shape. Significant color associations are associated with an asterisk. Where there was no significant color association for a letter, the most frequently associated color is depicted with a cross. There is no apparent systematic relationship between letter shape and color association.
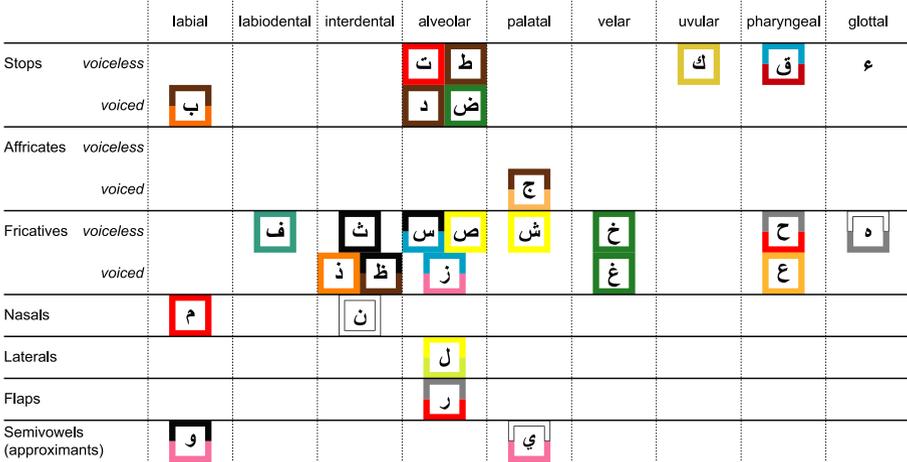
Figure 8. Letter sound and color associations. Letters are plotted according to their sound characteristics. Where there was no significant association for a letter, the most frequent color association is depicted. There are no apparent systematic relationships between letter sounds and color associations. This figure is published in color in the online version.

Kay (1969) hierarchy color terms enter into languages in the following order: (1) black/white; (2) red; (3) green/yellow; (4) blue; (5) brown; (6) orange/purple/grey/pink. Because only six stages are distinguished in this hierarchy, analyses were only performed for the first six letters of the alphabet. We computed correlations between the order of the color word responses as given by participants and the Berlin and Kay hierarchy, and ease of generation order (Al-Rasheed *et al.*, 2011). The correlations were tested separately for the different versions of the survey. Recall stimuli were presented either in random or fixed order, and starting with letter or number stimuli. These differences in stimulus order are relevant here because effects of ease of generation or the Berlin and Kay hierarchy could be triggered by any ordered sequence, or alternatively triggered by associations to intrinsic stimulus order (e.g., irrespective of when it is presented, the number 1 always becomes associated with the same color). Participants differed for each version of the survey. English and Arabic were also different from one another.

No effects were found for the ease of generation order of color words as found in Al-Rasheed *et al.* (2011). We did, however, find effects of the Berlin and Kay hierarchy (1969) summarized in Table 4. Table 4 shows color word responses to numbers followed the Berlin and Kay hierarchy in almost all cases (significant Pearson correlations, see Table 4), irrespective of the order in which stimuli were presented, irrespective of the response language, and irrespective of whether the survey started with numbers or letters. This means numbers positioned early in the intrinsic number sequence of 0–9 are associ-

Table 4.
Order effects in color associations*

	Correlation Berlin and Kay hierarchy with intrinsic stimulus order	Correlation Berlin and Kay hierarchy with random presentation order of stimuli
Random 1, letters first		
Letters Arabic ($N = 100$)	$p < 0.01, r = 0.260$	n.s.
Letters English ($N = 86$)	$p < 0.05, r = 0.254$	n.s.
Numbers E. Arabic ($N = 158$)	$p < 0.01, r = 0.250$	$p < 0.01, r = 0.230$
Numbers English ($N = 131$)	$p < 0.05, r = 0.206$	n.s.
Random 2, numbers first		
Letters E. Arabic ($N = 120$)	n.s.	n.s.
Letters English ($N = 46$)	$p = 0.068, r = 0.271$	n.s.
Numbers E. Arabic ($N = 212$)	$p < 0.0001, r = 0.280$	n.s.
Numbers English ($N = 87$)	n.s.	n.s.
Fixed 1, letters first		
Letters Arabic ($N = 117$)	n.s.	NA
Letters English ($N = 48$)	n.s.	NA
Numbers E. Arabic ($N = 186$)	$p < 0.001, r = 0.350$	NA
Numbers English ($N = 67$)	$p < 0.05, r = 0.279$	NA
Fixed 2, numbers first		
Letters Arabic ($N = 66$)	n.s.	NA
Letters English ($N = 36$)	n.s.	NA
Numbers E. Arabic ($N = 106$)	$p < 0.001, r = 0.482$	NA
Numbers English ($N = 58$)	$p < 0.001, r = 0.426$	NA

* For letter stimuli, the first six letters of the alphabet were used in the analyses. For numbers, all numbers were included ($N = 10$). Only stimuli belonging to Session 1 were analyzed. Reporting Pearson correlations. n.s. = nonsignificant ($p > 0.10$).

ated with color words early in the Berlin and Kay hierarchy (e.g., black, white, red), and this relationship is relatively robust.

For letter stimuli, the relationship of color word responses with the Berlin and Kay hierarchy is also present, but reaches significance only when stimuli were presented in random order (Random 1, letters first, see Table 4). The relationship between the order of letter stimuli in the alphabet and the Berlin and Kay hierarchy is therefore perhaps less strong than the relationship for numbers.

Finally, it is interesting to note the correlation between graphemes and the Berlin and Kay evolutionary sequence follows the intrinsic order of the items, and not the experimental order. That is, when the items are presented in a randomized order, participants did not associate the first letter with the first color in the evolutionary sequence. Rather 'alif ׀ was still more likely to be

Table 5.
Richness of color responses*

	Total responses	Rich responses	% Rich responses
Nonsynaesthetic responses	4970	482	9.70
Synaesthetic responses	1948	374	19.20

* Rich responses are nonstandard color words and multiword responses.

associated with 'black' or 'white', baa' ﺏ was more likely to be associated with 'red', etc., regardless of where in the experiment they occurred.

5.7. Richer Color Word Responses in Relation to Synaesthesia, Regardless of Gender

Previous studies report synaesthetes describe their color associations with more words and more nonstandard color words than nonsynaesthetes (Simner *et al.*, 2005). We found participants who displayed higher color consistency in their associations and therefore scored higher on the 'synaesthesia scale' tended to use more multiword responses (e.g., 'light green') or nonstandard color responses (e.g., 'silver'): Pearson correlation $r = 0.438$, $p < 0.001$. This analysis was collapsed across all participants, versions, and stimuli of the survey. The correlation is significant for both men and women (men $r = 0.510$, $p < 0.05$; women $r = 0.409$, $p < 0.001$). If we split all responses in 'synaesthetic' (matching between repetitions 1 and 2 of the questionnaire) and nonsynaesthetic (nonmatching responses), synaesthetic responses are roughly two times as likely as nonsynaesthetic responses to be a nonstandard ('exotic') color word or a multiword response (19.2% *versus* 9.70% of all responses, see Table 5).

6. Discussion

We found significant nonrandom pairings of graphemes and colors for both letter and number stimuli, amongst both Arabic and English responders (Figs 4 and 5), consistent with previous studies. However, the pairings themselves differ to those reported earlier. Table 6 summarizes grapheme–color associations reported for nonsynaesthetes in prior work, as a point of comparison. Despite the fact many studies have investigated letter–color associations in various languages, the primary associations are rarely reported in full. Therefore it is difficult to ascertain with certainty which letter–color association are cross-culturally robust. Table 6 shows there are considerable divergences between previously reported studies.

Table 6.

Reported grapheme–color associations for nonsynaesthetes*. This table is published in color in the online version

A: Latin alphabet

Dutch[1]

English[1–3]

German[4]

B: Devanāgarī

Hindi[1]

C: Arabic[5]

Egyptian Arabic

Table 6.
(Continued)

D: Numerals

	0	1	2	3	4	5	6	7	8	9
English[3]	□		■			■			■	
Eastern Arabic[5]	٠ □	١ ■	٢ ■	٣ ■	٤ ■	٥ ■	٦ ■	٧ ■	٨ ■	٩ ■

* Rich *et al.* (2005) and Simner *et al.* (2005) do not mention whether graphemes were tested in lowercase or uppercase; the first reports results using uppercase, the second using lowercase. In Rouw *et al.* (2014), all letters except n, b, l were presented in uppercase. Although nonsynesthaetes' grapheme–color associations have also been collected for the Chinese Hanzi and Japanese Hiragana, Katakana and Kanji scripts, the associations have not been reported. In all of the writing systems above, the relation between graphemes and phonemes is complex and involves one to many or even many to many mappings. Thus, while transliterations for the non-Latin alphabets are included, there is no guarantee that Hindi व /w, v/ is identical to English /w/ or Arabic و /w/, just as there is no guarantee that English 'w' and German 'w' or English 'a' and Dutch 'a' refer to the same sound. To avoid such misinterpretations, we have not attempted to align the sounds across the writing systems here.

1 Rouw *et al.* (2014).
2 Rich *et al.* (2005).
3 Spector and Maurer (2011).
4 Simner *et al.* (2005).
5 Current study.

For example, many studies find 'a' is mapped to red (e.g., Asano and Yoko-sawa, 2011; Guillamon, 2013; Marks, 1975; Rich *et al.*, 2005; Rouw *et al.*, 2014; Simner *et al.*, 2005). This has been found in closely related languages using the Roman script, like English, Dutch, and German, but also in languages with a different script, such as Hindi and Japanese. Some researchers suggest these mappings could be due to the shared sound. Guillamon (2013) has reported Arabic speakers also associate 'a' with red when mapping sounds to colors. In our study, however, all participants regardless of responding language (Arabic or English), strongly associated 'a' (which corresponds to the letter 'alif ‎ا‎) with black or white instead. There was no reliable association between 'a' and red.

Note, however, what may seem like considerable uniformity in previous studies masks some critical differences. The letter 'a' in Dutch and German is pronounced as the open front /a:/. But in English 'a' is pronounced as a midfront vowel /eɪ/ (as in 'age'). So, when English speakers associate the letter 'a' with red (e.g., Simner *et al.*, 2005) it is not entirely clear this is based on the specific sound. So it is unclear whether previously reported associations between 'a' and red are really based on the sound *per se*.

Rouw *et al.* (2014) suggest 'a' is red because it is the first (ordinal) item in a sequence, but we see in Arabic a very strong preference for the first ordinal item to be black or white. This is also reflected in the color associations to numbers. Participants in this study associated ٠ 0 and ١ 1 with black and white. Here it seems relevant that the letter 'alif ‎ا‎ and the number 1 ١ share the same shape. Previously, Beeli *et al.* (2007) and Smilek *et al.* (2007), found English-speaking synaesthetes associated 0 and 1 with black and white too, while Day (2004) found the letters I and O were associated with white or black by the majority of synaesthetes ($N = 166$). Baron-Cohen *et al.* (1993) found O to be associated with white, too, in 73% of synaesthetes. Rich *et al.* (2005) found I and O to be significantly associated with white in synaesthetes and I with white in nonsynaesthetes; Spector and Maurer (2008, 2011) also report preliterate children associate O and I with white. Together this might suggest a general shape–color association. The effect, however, appears to be strongest in synaesthetes. In contrast, Rouw *et al.* (2014) found Hindi nonsynaesthetes associated 1 with red. So this also suggests the posited universality of O and I shapes to black and white also deserves future scrutiny.

Within our own study the effect of shape appears to be limited. We hypothesized Arabic letter shape could play a role in determining associated colors, since many letters of the alphabet share similar shapes. However, letters with similar shapes were not assigned similar colors. In addition, no trends were detectable in color associations for letters with similar articulatory features.

We also tested whether the templatic morphology of Arabic has consequences for grapheme–color associations by checking whether color associations are driven by the onset letter of the raw color word, the root color morpheme, or both. We found a higher than chance tendency to match the stimulus letter to the onset of the raw color word and its root. However, the specific letter stimulus onset matches reached significance only in one case. This could be because there are two possible sources of association (raw letter onset; root letter onset), or perhaps because many Arabic color words start with the same letter, namely ١ ʾalif (Al-Jehani, 1990).

We found a clear positive relationship between the frequency of occurrence of letter and number stimuli in Arabic and the frequency of color words chosen as associates. That is, high frequency graphemes were associated with high frequency color words. There was also a relationship between letter frequency and the root token frequency of color words — a linguistic feature not previously examined. The relationship between letter frequency and word frequency has previously been reported by Simner *et al.* (2005), but only for synaesthetes. Here we found the effect for nonsynaesthetes too. Simner *et al.*'s study features a large number of nonsynaesthete participants ($N = 317$ for English- and $N = 58$ for German-speaking participants), so it is unlikely the discrepancy between the results is due to a lack of power in the Simner *et al.* study. In addition, Simner *et al.* did not correct for multiple comparisons (which we did), making it even more unlikely an effect was missed by Simner *et al.* So, why this difference? The discrepancy could be driven by language-specific factors. Perhaps the relative frequencies of letters in Arabic differ more than in English or German. However, this requires further in-depth studies to verify or refute.

In our data, the order in which color word responses were generated did not follow the ease of generation order for Arabic speakers as reported by Al-Rasheed *et al.* (2011). This ease of generation order effect was previously reported for nonsynaesthetes by Simner *et al.* (2005). One reason we did not find this effect could be that our source literature was not specific enough for the participants of our study, who originated mainly from Egypt. The Al-Rasheed *et al.* color word generation norms were from Arabic speakers from Saudi Arabia. Separate ease of generation norms are required from Egyptian Arabic speakers to further verify this. In our data, order of color word responses for number stimuli correlated strongly with the Berlin and Kay evolutionary sequence, even when the stimuli were presented in random order. For letter stimuli, this effect was also found, contra Simner *et al.* (2005) who reported no effect. This relationship may be stronger for number stimuli because of the stronger intrinsic ordering of number *vs.* letter stimuli (Shanon, 1982).

6.1. Synaesthetes in our Study

By means of our cutoff method, we identified two synaesthetes in our study, out of 78 participants who completed both sessions of the experiment (2.6%). The use of a statistical cutoff naturally implies that a certain percentage of participants would fall above the specified value. As such it is not possible to make definitive claims about the prevalence of synaesthesia in our sample. Our purpose was merely to identify those participants exhibiting clear synaesthetic characteristics. Comparisons to earlier prevalence studies on populations using the Roman script (Rothen and Meier, 2010; Simner *et al.*, 2006) suggest that our method yields a rather conservative estimate of synesthesia prevalence.

Our results confirm the tendency of synaesthetes to describe their specific color associations in more detail than regular participants. For instance, Simner *et al.* (2005) found 70 synaesthete participants together generated 54 different color descriptions for the color green, compared to only five different descriptions of green from control participants. In our data, responses matching across the first and second session of the survey (and thus regarded as 'synaesthetic' in nature) consisted of multiword utterances or 'exotic' color words more often than nonmatching responses. This indicates that for certain participants color associations may have been rather vivid or akin to color imagery, allowing specific color associations to occur. Tentatively, the detail with which participants systematically describe their color associations might even be a possible way to identify synaesthetes.

7. Conclusion

In recent years there is growing acknowledgement that psychology needs to expand its investigations to a more representative sample of humanity (Henrich *et al.*, 2010; Majid and Levinson, 2010). We find color associations to letters and numbers in an Arabic speaking population differ in many respects from previous results for speakers of other languages, such as English. In fact, the results suggest grapheme–color associations are determined to a large extent by language-internal psycholinguistic factors such as letter frequency, the intrinsic ordering of letters, and the morphological structure of Arabic. Other features previously suggested as important, such as letter shape and sound, appear to play a negligible role. Grapheme–color associations appear to be the result of statistical associations derived from the particular language and cultural context speakers find themselves in. The findings together show linguistic and psycholinguistic features have to be carefully examined in each language individually in order to provide a pancultural account of grapheme–color associations.

Notes

1. http://www.intellaren.com/articles/en/a-study-of-arabic-letter-frequency-analysis.

2. The degrees of freedom vary between these two analyses because Aralex did not have information for all root token frequencies.

References

Al-Jehani, N. M. (1990). Color terms in Mecca: a sociolinguistic perspective, *Anthropol. Linguist.* **32**, 163–174.

Al-Rasheed, A. S., Al-Sharif, H. H., Thabit, M. J., Al-Mohimeed, N. S. and Davies, I. R. L. (2011). Basic-color terms of Arabic, in: *New Directions in Colour Studies*, C. P. Biggam, C. A. Hough, C. J. Kay and D. R. Simmons (Eds), pp. 53–58. John Benjamins Publishing Company, Amsterdam, The Netherlands.

Asano, M. and Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes, *Conscious. Cogn.* **20**, 1816–1823.

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

Baron-Cohen, S., Harrison, J., Goldstein, L. H. and Wyke, M. (1993). Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception* **22**, 419–426.

Battig, W. F. and Montague, W. E. (1969). Category norms for verbal items in 56 categories — a replication and extension of Connecticut category norms, *J. Exp. Psychol.* **80**, 1–45.

Beeli, G., Esslen, M. and Jancke, L. (2007). Frequency correlates in grapheme–color synaesthesia, *Psychol. Sci.* **18**, 788–792.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA, USA.

Boudelaa, S. and Marslen-Wilson, W. D. (2010). Aralex: a lexical database for modern standard Arabic, *Behav. Res. Methods* **42**, 481–487.

Brang, D., Rouw, R., Ramachandran, V. S. and Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme–color synesthesia, *Neuropsychologia* **49**, 1355–1358.

Buckwalter, T. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0 LDC2002L49, in: Linguistic Data Consortium, Philadelphia, PA, USA.

Campbell, G. L. (1997). *Handbook of Scripts and Alphabets*. Routledge, New York, NY, USA.

Cohen Kadosh, R., Henik, A. and Walsh, V. (2007). Small is bright and big is dark in synaesthesia, *Curr. Biol.* **17**, R834–R835.

Cohen Kadosh, R., Cohen Kadosh, K. and Henik, A. (2008). When brightness counts: the neuronal correlate of numerical–luminance interference, *Cereb. Cortex* **18**, 337–343.

Coulmas, F. (2013). *Writing and Society: an Introduction*. Cambridge University Press, Cambridge, MA, USA.

Day, S. A. (2004). Trends in Synesthetically Colored Graphemes and Phonemes — 2004 revision. Available from http://www.daysyn.com/Day2004Trends.pdf.

Deroy, O. and Spence, C. (2013). Why we are not all synesthetes (not even weakly so), *Psychon. Bull. Rev.* **20**, 643–664.

Elbeheri, G. and Everatt, J. (2006). Literacy ability and phonological processing skills amongst dyslexic and non-dyslexic speakers of Arabic, *Read. Writ.* **20**, 273–294.

Foster, G. M. (1979). Methodological problems in the study of intracultural variation: the hot/cold dichotomy in Tzintzuntzan, *Hum. Organ.* **38**, 179–183.

Guillamon, P. M. (2013). Vowel–colour symbolism in English and Arabic: a comparative study, *Misc. J. Engl. Am. Stud.* **47**, 31–52.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010). The weirdest people in the world, *Behav. Brain Sci.* **33**, 1–75.

Hochel, M. and Milán, E. G. (2008). Synaesthesia: the existing state of affairs, *Cogn. Neuropsychol.* **25**, 93–117.

Hung, W.-Y., Simner, J., Shillcock, R. and Eagleman, D. M. (2014). Synaesthesia in Chinese characters: the role of radical function and position, *Conscious. Cogn.* **24**, 38–48.

Jürgens, U. M. and Nikolić, D. (2012). Ideaesthesia: conceptual processes assign similar colours to similar shapes, *Transl. Neurosci.* **3**, 22–27.

Kusnir, F. and Thut, G. (2012). Formation of automatic letter–colour associations in non-synaesthetes through likelihood manipulation of letter–colour pairings, *Neuropsychologia* **50**, 3641–3652.

Lewis, P. M., Simons, G. F. and Fennig, C. D. (2014). *Ethnologue: Languages of the World*, 17th edn. SIL International, Dallas, TX, USA.

Majid, A. and Levinson, S. C. (2010). WEIRD languages have misled us, too, *Behav. Brain Sci.* **33**, 103.

Malt, B. C. and Majid, A. (2013). How thought is mapped into words, *Wiley Interdiscip. Rev. Cogn. Sci.* **4**, 583–597.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions, *Psychol. Bull.* **82**, 303–331.

Rich, A. N., Bradshaw, J. L. and Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical–colour associations, *Cognition* **98**, 53–84.

Rothen, N. and Meier, B. (2010). Higher prevalence of synaesthesia in art students, *Perception* **39**, 718–720.

Rouw, R., Gosavi, R., Case, L. and Ramachandran, V. (2014). Color associations for days and letters across different languages, *Front. Psychol.* **5**, 369. DOI:10.3389/fpsyg.2014.00369.

Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press, Cambridge, UK.

Shanon, B. (1982). Colour associates to semantic linear orders, *Psychol. Res.* **44**, 75–83.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L. and Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations, *Cogn. Neuropsychol.* **22**, 1069–1085.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K. and Ward, J. (2006). Synaesthesia: the prevalence of atypical cross-modal experiences, *Perception* **35**, 1024–1033.

Simner, J., Hung, W. Y. and Shillcock, R. (2011). Synaesthesia in a logographic language: the colouring of Chinese characters and Pinyin/Bopomo spellings, *Conscious. Cogn.* **20**, 1376–1392.

Smilek, D., Carriere, J. S. A., Dixon, M. J. and Merikle, P. M. (2007). Grapheme frequency and color luminance in grapheme–color synaesthesia, *Psychol. Sci.* **18**, 793–795.

Spector, F. and Maurer, D. (2008). The colour of Os: naturally biased associations between shape and colour, *Perception* **37**, 841–847.

Spector, F. and Maurer, D. (2011). The colors of the alphabet: naturally-biased associations between shape and color, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 484–495.

Ward, J. and Simner, J. (2003). Lexical–gustatory synaesthesia: linguistic and conceptual factors, *Cognition* **89**, 237–261.

Ward, J., Huckstep, B. and Tsakanikos, E. (2006). Sound–colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex* **42**, 264–280.

Watson, M. R., Akins, K. and Enns, J. T. (2010). Similar letters, similar hues — shape–colour isomorphism in grapheme–colour synaesthesia, in: *2010 Meeting of the UK Synaesthesia Association*, Brighton, UK.

Watson, M., Akins, K. and Enns, J. (2012). Second-order mappings in grapheme–color synesthesia, *Psychon. Bull. Rev.* **19**, 211–217.

Wollen, K. A. and Ruggiero, F. T. (1983). Colored-letter synesthesia, *J. Ment. Imagery* **7**, 83–86.