UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Do the colors of your letters depend on your language? Language-dependent and universal influences on grapheme-color synesthesia in seven languages

Root, N.; Asano, M.; Melero, H.; Kim, C.-Y.; Sidoroff-Dorso, A.V.; Vatakis, A.; Yokosawa, K.; Ramachandran, V.; Rouw, R.

DOI
10.1016/j.concog.2021.103192

Publication date
2021

Document Version
Final published version

Published in
Consciousness and Cognition

License
CC BY

Link to publication

Citation for published version (APA):
Root, N., Asano, M., Melero, H., Kim, C.-Y., Sidoroff-Dorso, A. V., Vatakis, A., Yokosawa, K., Ramachandran, V., & Rouw, R. (2021). Do the colors of your letters depend on your language? Language-dependent and universal influences on grapheme-color synesthesia in seven languages. *Consciousness and Cognition*, *95*, Article 103192. https://doi.org/10.1016/j.concog.2021.103192

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Sept 2026

Consciousness and Cognition 95 (2021) 103192

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Do the colors of your letters depend on your language? Language-dependent and universal influences on grapheme-color synesthesia in seven languages

Nicholas Root [a,*], Michiko Asano [b,1], Helena Melero [c,1], Chai-Youn Kim [d,1], Anton V. Sidoroff-Dorso [e,f,1], Argiro Vatakis [g,1], Kazuhiko Yokosawa [h], Vilayanur Ramachandran [i], Romke Rouw [a]

[a] Brain and Cognition, Psychology Department, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, Netherlands
[b] Department of Psychology, College of Contemporary Psychology, Rikkyo University, 1-2-26 Kitano, Niiza, Saitama 352-8558, Japan
[c] Department of Psychobiology and Methodology in Behavioral Sciences, Faculty of Psychology, Complutense University of Madrid, Av. Séneca, 2, 28040 Madrid, Spain
[d] Department of Psychology, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
[e] Department of Psychology, Moscow Pedagogical State University, Moscow 119435, Russia
[f] Centre of Interdisciplinary Research of Contemporary Childhood, Moscow State University of Psychology and Education, Moscow 123290, Russia
[g] Department of Psychology, Panteion University of Social and Political Sciences, 136 Syngrou Ave., 17671 Athens, Greece
[h] Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[i] Department of Psychology, University of California, San Diego, 9500 Gilman Drive, San Diego, CA 92093, United States

ARTICLE INFO

Keywords:
Synesthesia
Grapheme-color
Cross-linguistic
Multisensory
Regulatory factor

ABSTRACT

Grapheme-color synesthetes experience graphemes as having a consistent color (e.g., "N is turquoise"). Synesthetes' specific associations (*which* letter is *which* color) are often influenced by linguistic properties such as phonetic similarity, color terms ("Y is yellow"), and semantic associations ("D is for dog and dogs are brown"). However, most studies of synesthesia use only English-speaking synesthetes. Here, we measure the effect of color terms, semantic associations, and non-linguistic shape-color associations on synesthetic associations in Dutch, English, Greek, Japanese, Korean, Russian, and Spanish. The effect size of linguistic influences (color terms, semantic associations) differed significantly between languages. In contrast, the effect size of non-linguistic influences (shape-color associations), which we predicted to be universal, indeed did not differ between languages. We conclude that language matters (outcomes are influenced by the synesthete's language) and that synesthesia offers an exceptional opportunity to study influences on letter representations in different languages.

1. Introduction

Grapheme-color synesthesia is a phenomenon in which graphemes are experienced as having a consistent color (e.g., "The letter N

* Corresponding authors.
E-mail address: n.root@uva.nl (N. Root).
[1] Authors contributed equally.

https://doi.org/10.1016/j.concog.2021.103192
Received 10 November 2020; Received in revised form 5 August 2021; Accepted 6 August 2021
Available online 6 September 2021
1053-8100/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

is saffron orange"). One defining property of synesthesia is the qualia-laden perceptual nature of synesthetes' experience: there is converging evidence from behavioral (e.g., Ramachandran & Hubbard, 2001; Palmeri, Blake, Marois, Flanery, & Whetsell, 2002; for a review see Kim & Blake, 2013), electrophysiological (Brang, Hubbard, Coulson, Huang, & Ramachandran, 2010), and fMRI (Nunn et al., 2002; Hubbard et al., 2005; but see Hupé et al., 2012; Melero et al., 2014) studies that synesthetes *actually see* their synesthetic colors. Another defining property of synesthesia is that the particular associations of any one synesthete (*which* letter is *which* color) are highly consistent over time: when asked to match graphemes to their colors, a synesthete will pick the same color for each grapheme even when re-tested after months or years (Asher, Aitken, Farooqi, Kurmani, & Baron-Cohen, 2006; but see Simner, Ipser, Smees, & Alvarez, 2017; Chromý et al., 2019). The colors of synesthetes' graphemes were once thought to be idiosyncratic, but research using large synesthesia datasets has conclusively demonstrated that certain graphemes are more frequently associated with certain colors (Rich et al., 2005; Simner et al., 2005). Intriguingly, similar non-random distributions of grapheme-color combinations are found in non-synesthetes when they are forced to generate color associations for graphemes (Rouw, Case, Gosavi & Ramachandran, 2014; Simner et al., 2005; van Leeuwen et al., 2016), suggesting that these associations may arise from cognitive or developmental processes that are not specific to synesthetes.

Why are some grapheme-color combinations more probable than others? One possibility is that statistical regularities or meaningful associations in the real world influence synesthetic colors during development (e.g., Newell & Mitchell, 2016). For example, color terms strongly influence synesthetic associations: "R" is often red, "Y" is often yellow, and so on (Rich et al., 2005; Simner et al., 2005). More subtly, female synesthetes tend to associate their first initial with the color pink (Root, Dobkins, Ramachandran, & Rouw, 2019), consistent with gender stereotypes prevalent during childhood (the period during which synesthetic associations develop; Simner et al., 2009). These are just two examples of more than a dozen different influences on synesthetic association. Different authors have used different terms to describe some of these influences: "first- and second-order mappings" (Watson et al., 2012), "cross-modal correspondences" (Spector & Maurer, 2011), "underlying mechanisms" (Simner et al., 2005), and "Regulatory Factors" (Asano et al., 2017). In the present work, we choose to use the term "Regulatory Factor" (RF) because we believe it is the broadest, most inclusive term for the phenomenon we wish to describe. We define an RF as any rule that makes a set of predictions about the observed pattern of grapheme-color associations. By this definition, first- and second-order mappings (Watson et al., 2011) are both *types* of RFs, and the superordinate term "mapping" is still not general enough to describe some phenomena we would call RFs. For example, the tendency for synesthetes to associate letters *in general* with warmer colors (the synesthetic "palette"; Rouw & Root, 2019) is an RF that explains a substantial amount of variance in synesthetes' color associations, but it is not a "mapping".

Many of the RFs that influence grapheme-color associations are linguistic in origin (word associations, pronunciation, letter frequency, etc.); indeed, Simner (2007) suggests that grapheme-color synesthesia should be considered a psycholinguistic phenomenon as much as a perceptual one. This suggests that psycholinguistics can be used to understand synesthesia, and perhaps more importantly, that synesthesia can be used to understand psycholinguistics in the non-synesthetic population (Mankin, 2019; Simner, 2007). If many RFs are linguistic in origin, one obvious question is whether they are generally language-specific or language-universal. Unfortunately, almost all existing studies of grapheme-color synesthesia examined only English-speaking synesthetes. The few existing studies of synesthesia in other languages suggest that both types of RF might exist. At least one RF is plausibly universal: Root et al. (2018) found that the grapheme in the first ordinal position tends to be red in English, Spanish, Dutch, Japanese, and Korean. However, at least one RF might be language-specific: grapheme pronunciation influences grapheme-color associations in native speakers of Japanese (Asano & Yokosawa, 2011) and Korean (Kang, Kim, Shin, & Kim, 2017), but *not* in native English speakers (Watson, Akins, & Enns, 2012). Although these results are highly suggestive, the studies are not directly comparable, because pronunciation was operationalized in three *very* different ways: in Watson et al. (2012), presence of a shared vowel in the English alphabet spelling (b – "bee" – and g – "gee" – are similar); in Asano and Yokosawa (2011), presence of a shared consonant in the Japanese kana syllable ("ma" and "mi" are similar); in Kang et al. (2017), presence of a shared articulatory feature (the bilabial consonants "m" and "b" are similar). Thus, it is still not known whether RFs that are known to influence grapheme-color synesthesia in one language play the same role in other languages.

The present paper aims to show that if we want to understand where grapheme-color associations come from, "language matters". We find that while some RFs are plausibly universal, other RFs are stronger, weaker, or absent altogether in some languages. In one and the same dataset, we measure the influence of three previously reported RFs in synesthetes who are native speakers of Dutch, English, Greek, Japanese, Korean, Russian, and Spanish. We quantify the effect of the three RFs *using the same methodology*, so that effect sizes are directly comparable across language.

2. International dataset

2.1. Datasets

Dutch. 194 native speakers of Dutch were recruited from the general public (via television and radio interviews) and from the undergraduate participant pool at the University of Amsterdam. Demographic information was incomplete as subjects could opt out from reporting these, but most subjects were females 18–36 years old. All subjects completed the Eagleman Synesthesia Battery (Eagleman et al., 2007) in English (all subjects were fluent in English). Subjects were not exclusively self-reported synesthetes; subjects who were unsure of their synesthesia status were also encouraged to take the test.

English. 165 native speakers of English were recruited through fliers posted on the University of California, San Diego campus, as well as similar ads on the web. Demographic information was incomplete as subjects could opt out from reporting these, but most subjects were females 18–25 years old. All subjects completed the Eagleman Synesthesia Battery (Eagleman et al., 2007). Subjects were not exclusively self-reported synesthetes; non-synesthetes and subjects who were unsure of their synesthesia status were also

encouraged to take the test.

Greek. 23 native speakers of Greek who self-reported as synesthetes were recruited from students at the Department of Psychology at Panteion University, Athens, and also from the general public via advertisements in social media (e.g., Facebook, Twitter). Demographic information was incomplete as subjects could opt out from reporting these, but almost all subjects were females 19–47 years old. All subjects completed the Eagleman Synesthesia Battery (Eagleman et al., 2007), which was translated into Greek.

Japanese. 27 native speakers of Japanese (3 males, 24 females; mean age 23.81, range 18–44) who self-reported as synesthetes were recruited via a website (see Asano & Yokosawa, 2011). Japanese synesthetes selected a color experienced for each of the 46 basic Hiragana characters from a palette of the 138 named W3C colors (for details, see Supplemental Text S1 in Root et al., 2018).

Korean. 13 native speakers of Korean who self-reported as synesthetes were identified using the Korean Synesthesia Questionnaire (see Shin & Kim, 2014). Demographic information was incomplete, but most subjects were females 21–27 years old. Data for eight subjects was acquired using a translated version of the synesthesia test derived from the TexSyn Toolbox for Matlab, that was functionally identical to the Synesthesia Battery (Eagleman et al., 2007); data for the other five subjects was acquired by asking synesthetes to adjust the color of a square to match each inducing grapheme, using the color palette embedded in Microsoft Powerpoint.

Russian. 22 native speakers of Russian who self-reported as synesthetes were identified using public announcements in an internet-based social networking forum (Russian Synaesthesia Community) by author AVSD. Demographic information was incomplete, but most subjects were females 18-56 years old. All subjects completed the Eagleman Synesthesia Battery (Eagleman et al., 2007), which was translated into Russian.

Spanish. 62 native speakers of Spanish were recruited using the Artecittá Foundation Questionnaire (see Melero, Peña-Melián, & Rios-Lago, 2015). Demographic information was incomplete as subjects could opt out from reporting these, but almost all subjects were females 18–36 years old. All subjects completed the Eagleman Synesthesia Battery (Eagleman et al., 2007), which was translated into Spanish. Subjects were not exclusively self-reported synesthetes; non-synesthetes and subjects who were unsure of their synesthesia status were also encouraged to take the test.

Datasets from Dutch, English, Japanese, Korean, and Spanish include subjects that were reported in a previous international collaboration (Root et al., 2018). In the present work, we use the raw unfiltered data (even for the datasets that were preprocessed in a previous study) so that we can match the preprocessing steps between languages as closely as possible.

For all languages, only native speakers of the language were included, but we did not restrict our study to monolinguals. All Dutch subjects were fluent in English. All Greek subjects had at least B2 levels of English proficiency. Although we do not have second-language proficiency data for all Japanese subjects, English is taught as part of compulsory education in Japanese junior high schools, and 70% of the subjects (19/27) were university students in and around Tokyo, who tend to have mid to high levels of English proficiency. Similarly, most Korean subjects studied English as their second language (for as many as 10 years at the time of data collection). We do not have second-language proficiency data for Spanish subjects, but most of these subjects were university students in Madrid, who must take an English proficiency test as part of their college entrance exam. Almost all Russian subjects have only elementary command of other languages (indeed, only 11% of Russian native speakers are fluent in English; Владение ИностранныМи ЯзыкаМи, 2014). In sum, although we do not have demographic data on second language proficiency for every subject, we can safely assume that most of the subjects in our study (other than Russian subjects) have some degree of English-language fluency. In the countries in our study, English second language instruction typically begins around age 10, after many synesthetic associations have already been "locked in" (Simner & Bain, 2013). In Supplemental Text S2, we directly test whether English proficiency could have influenced grapheme-color associations in our non-English datasets; we find no evidence of such influence.

2.2. Preprocessing

We first removed from our analyses subjects who did not report colors for at least 50% of the graphemes in their language (either because they did not experience colors for these graphemes, or because their batteries were incomplete). Color values for all trials were converted to the perceptual CIELuv color space, which has been shown to maximize sensitivity and specificity for diagnosis of synesthesia using test–retest consistency (Rothen, Seth, Witzel, & Ward, 2013). For the purposes of evaluating test–retest consistency, we considered only graphemes in which subjects chose a color on all three trials (on the Eagleman battery, it is possible to indicate "no color"). Synesthesia was operationalized using the criterion of Rothen et al. (2013): the test–retest consistency for each grapheme was calculated as the sum of Euclidean distances between the three repeated trials, and subjects were classified as synesthetes if the average test–retest consistency of their graphemes was <135. We also added an additional constraint to Rothen's criterion: subjects only qualified as synesthetes if the test–retest consistency of 1000 randomly-shuffled permutations of their data was significantly higher ($>95\%$ of permutations) than the 135 cutoff. This additional criterion was designed to exclude subjects who chose the same color repeatedly throughout the experiment; these subjects chose black or very dark grey on almost every trial, and are plausibly non-synesthetes who misunderstood task instructions and chose the color in which the grapheme was printed on their screen (black).

After removing subjects who did not meet these criteria, our final dataset contained 118 Dutch, 57 English, 15 Greek, 27 Japanese, 8 Korean, 17 Russian, and 21 Spanish synesthetes.

Next, the three test–retest repeats for each grapheme (for each subject) were collapsed into a single value by taking the average (in CIELuv color space) of the repeated measurements for each grapheme[2]. If this average was further than 45 CIELuv units from one of the trials, the trial was removed, and the average was recomputed. This prevented the averaging process from producing average colors that are unrelated to any colors reported by the subject. For example, if a synesthete reports on three test–retest trials that a grapheme is bright red, bright green, and an identical bright green, the CIELuv average of these trials is a desaturated yellow; our method excludes the red trial, yielding a new CIELuv average of bright green. In addition, if a synesthete reports on three test–retest trials that a grapheme is bright red, bright green, and bright blue, this would average to white, but our method instead excludes all three trials, and thus excludes the grapheme from analysis.

The hypotheses tested in this paper are concerned with the *categories* of color experienced by our subjects (e.g., the propensity to associate a grapheme with a red color, rather than a *particular* shade of red). We therefore categorized our synesthetes' color associations into the 11 basic colors of Berlin and Kay (1991). The averaged CIELuv coordinates were classified into Berlin-Kay color terms using the Colournamer algorithm of Mylonas, MacDonald, and Wuerger (2010), a synthetic observer trained using color-naming data of thousands of human subjects in many different languages. Since color term boundaries can differ between language (e.g., Coventry et al., 2006; Paramei et al., 2018; Regier et al., 2005), we had to choose between using language-specific color boundaries for each language or English color boundaries for all the languages. We chose to use English color boundaries for all languages in our primary analysis, so that the same color would be categorized identically across language. To ensure that this is not a confound, we replicated our analyses using language-specific color boundaries and found that our results are the same (our choice to use English-language color boundaries did not matter); see Supplemental Material. For the analyses presented in the main manuscript, the averaged CIELuv coordinates for each association were classified as the English-language Berlin-Kay color with the highest posterior probability in the Colournamer algorithm.

3. Regulatory Factor 1: *Color Term RF* – "Y is yellow"

The influence of color terms on synesthetic color associations is perhaps the most intuitive and obvious regulatory factor that has been described in the literature: for English-speaking synesthetes, "R" is often red, "Y" is often yellow, etc. This observation was first noted qualitatively in Day (2004), and was subsequently reported in Rich et al. (2005) and tested statistically in Simner et al. (2005). In Simner et al.'s study, synesthetes report the name of the color they experience (e.g., "crimson"), and an association is classified as a "match" if the initial letter of the color was the same as the grapheme it was associated with (e.g., if a synesthete said that "C is crimson"). Simner et al. (2005) find approximately twice as many matches as would be expected by chance, yielding strong support for an influence of color names on synesthetic associations. Here, we attempt to replicate the Color Term RF in all seven languages in our dataset. Since synesthetes in our dataset used a colorpicker to report their color associations (rather than reporting the name of the color), and since non-basic color categorization can differ significantly between languages (e.g., Paramei et al., 2018), we chose to restrict our test of the Color Term RF to the names of the 11 basic colors of Berlin and Kay (1991) in each language. Specifically, we test the hypothesis that the initial letter of each Berlin-Kay color term is associated with that color ("R" is red, "Y" is yellow, etc.).

3.1. Methods

Table 1 lists the Berlin-Kay colors for each language in our dataset. Note that in Russian and Greek, there is an obligatory distinction between light blue (голубой and γαλάζιο, respectively) and dark blue (синий and μπλε, respectively); in other words, these languages technically have 12 Berlin-Kay colors rather than 11 (Coventry et al., 2006; Davies & Corbett, 1994; Winawer et al., 2007). In our main analysis, we chose to count any association with blue as a match for these letters, in order to keep our color categorization method consistent across languages. In the Supplemental Material, we report the results of the analysis using Russian and Greek color boundaries. The choice to merge blues did not affect our results: no significant result became non-significant, or vice versa.

3.2. Results

For each language, we computed $matches_{observed}$, the number of grapheme-color associations that were consistent with the predictions of the Color Term RF. We also computed $matches_{possible}$, the number of grapheme-color associations that *could* have been consistent, by counting the total number of reported associations that were the initial letter of a Berlin-Kay term. The value of $matches_{possible}$ for a language is influenced both by the idiosyncrasies of the color terms in a language (e.g., if there are multiple color terms that begin with the same grapheme, $matches_{possible}$ will be smaller), and also by the idiosyncrasies of individual subjects'

[2] Preprocessing for Japanese synesthetes was different than for other languages, because Japanese synesthetes had completed a color-picker battery with two repeats instead of three, making our trial exclusion and averaging procedure impossible. Japanese synesthetes were classified as synesthetes using the same threshold as the other languages (average Euclidean distance in CIELuv between the two repeated trials$<$45, i.e., 135/3); all 27 synesthetes met this criterion, but rather than averaging the repeats, data from the first repeat was used for subsequent analyses. For more details about this data (including an analysis of a subset of subjects demonstrating that using two *vs.* three repeats does not alter results), see Supplemental Text S1 in Root et al. (2018).

Table 1
Berlin-Kay color terms for each of the seven languages in the dataset. The initial grapheme (the grapheme predicted to be influenced by the Color Term RF) is in parenthesis.

English	Dutch	Greek	Japanese[1]	Korean	Russian	Spanish
Black (B)	Zwart (Z)	Μαύρο (Μ)	黒 (く)	검정 (ㄱ)	Черный (Ч)	Negro (N)
White (W)	Wit (W)	Άσπρο (Α)	白 (し)	흰색 (ㅎ)	Белый (Б)	Blanco (B)
Grey (G)	Grijs (G)	Γκρι (Γ)	灰色 (は)	회색 (ㅎ)	Серый (С)	Gris (G)
Red (R)	Rood (R)	Κόκκινο (Κ)	赤 (あ)	빨강 (ㅂ)	Красный (К)	Rojo (R)
Green (G)	Groen (G)	Πράσινο (Π)	緑 (み)	초록 (ㅊ)	Зеленый (З)	Verde (V)
Blue (B)	Blauw (B)	Γαλάζιο/Μπλε (Γ/Μ)	青 (あ)	파랑 (ㅍ)	Голубой/Синий (Г/С)	Azul (A)
Yellow (Y)	Geel (G)	Κίτρινο (Κ)	黄色 (き)	노랑 (ㄴ)	Желтый (Ж)	Amarillo (A)
Orange (O)	Oranje (O)	Πορτοκαλί (Π)	オレンジ (お)	주황 (ㅈ)	ОранЖевый (О)	Naranja (N)
Brown (B)	Bruin (B)	Καφέ (Κ)	茶色 (ち)	밤색 (ㅂ)	Коричневый (К)	Marrón (M)
Pink (P)	Roze (R)	Ροζ (Ρ)	ピンク (ぴ)[2]	분홍 (ㅂ)	Розовый (Р)	Rosa (R)
Purple (P)	Paars (P)	Μωβ (Μ)	紫 (む)	보라 (ㅂ)	Фиолетовый (Ф)	Morado (M)

[1] In the table, we list the most common written form for the color, and the initial kana of the Hiragana spelling.

[2] The initial grapheme for pink in Hiragana is ぴ, which is ひ plus a diacritical mark that changes voicing. For Japanese synesthetes – like synesthetes in other languages with diacritical marks – adding a diacritical mark can change the reported color of the grapheme, typically by making it brighter/darker or more/less saturated. In our analysis, we use the colors reported for ひ without its diacritic, but we verified that this did not influence our analysis: an equal number of synesthetes (n = 5) associated ひ *vs.* ぴ with pink.

associations (e.g., if a subject does not experience a color for a grapheme, $matches_{possible}$ will be smaller). The ratio $\frac{matches_{observed}}{matches_{possible}}$ is a "prediction-realization"-type pseudo-R^2 value (Veall & Zimmermann, 1992; Veall & Zimmerman, 1996) that represents the maximum proportion of associations in our sample that can be attributed to the Color Term RF. For the Color Term RF, this proportion varies from 19% (Spanish) to 42% (English).

Note that this pseudo-R^2 is the *maximum* proportion of associations that might be explained by the Color Term RF. Even if the Color Term RF exerts no influence on synesthetic associations, some of these associations will be consistent with the Color Term RF by chance, so the value of pseudo-R^2 under the null hypothesis is non-zero. For example, if each of the 11 basic colors had a different initial, and if color associations were random and uniformly distributed (p = 1/11 for each color), the expected value of pseudo-R^2 under the null hypothesis is $E[R^2|H_0] = 0.091$ (i.e., 9%). In our actual data, the expected value of pseudo-R^2 under H_0 is influenced in a complex way by the distribution of color terms in the alphabet (e.g., since black, blue, and brown all begin with "B", the probability of a match for "B" under the null hypothesis is much higher), and by the distribution of synesthetic colors for all letters (e.g., "Y" will often be yellow even under H_0 because synesthetic associations *in general* are often yellow; Rouw & Root, 2019).

The use of this particular pseudo-R^2 value ($\frac{matches_{observed}}{matches_{possible}}$) is common in the synesthesia literature (e.g., Mankin & Simner 2017; Root et al., 2018; Simner et al., 2005), but past work has estimated the expected value under the null ($E[R^2|H_0]$) using resampling/randomization tests. Here, we derive a general formula that gives the exact value of $E[R^2|H_0]$ for any RF that makes a prediction that certain graphemes should be associated with certain colors. Under the null hypothesis, in a dataset of size n, the expected number of times that a grapheme g is associated with a color c is $n \times P(g) \times P(c)$: the total number of associations in the data, multiplied by the proportion of associations that are the grapheme g,[3] multiplied by the proportion of associations that are the color c. The total expected number of matches under the null hypothesis for an RF that makes i predictions is the sum of the expected number of matches for each prediction, $E[matches_{observed}|H_0] = n\sum_{i \in RF} P(g_i)P(c_i)$. Our desired statistic $E[R^2|H_0]$ (the expected proportion of matches for the RF) is the expected number of matches divided by the possible number of matches:

$$E[R^2|H_0] = \frac{n\sum_{i \in RF} P(g_i)P(c_i)}{matches_{possible}}$$

From these values, we can assess the statistical significance of our pseudo-R^2 using a binomial test of the hypothesis that $matches_{observed}$ successes from $matches_{possible}$ trials is consistent with the expected proportion of matches $E[R^2|H_0]$. Fig. 1a depicts the expected pseudo-R^2, observed pseudo-R^2, and confidence intervals for each language; Table 2 (left side) lists these values, as well as the p-values for each binomial test. The proportion of associations consistent with the Color Term RF was higher than would be predicted by chance in Dutch, English, Japanese, Korean, and Russian, was trending ($p = 0.061$) in Greek, but was not significantly higher than would be predicted by chance ($p = 0.177$) in Spanish.

Although this pseudo-R^2 has an intuitive interpretation and it is straightforward to evaluate the statistical significance of an RF in a particular language, our pseudo-R^2 cannot be used to compare an RF between two languages for which $E[R^2|H_0]$ differs. To understand why this is true, consider a hypothetical language in which the pseudo-R^2 is large but close to the expected value under the null (e.g., $R^2 = 0.8$ and $E[R^2|H_0] = 0.79$) and a hypothetical language in which the pseudo-R^2 is small but far from the expected value under the null (e.g., $R^2 = 0.4$ and $E[R^2|H_0] = 0.01$). It does not make sense to claim that the Color Term RF is "twice as strong" in the first language, since a high value of pseudo-R^2 would be expected by chance; indeed, it seems more reasonable to say that the Color Term RF

[3] Note that $P(g)$ is not 1/26, because not every synesthete experiences a color for all 26 letters.

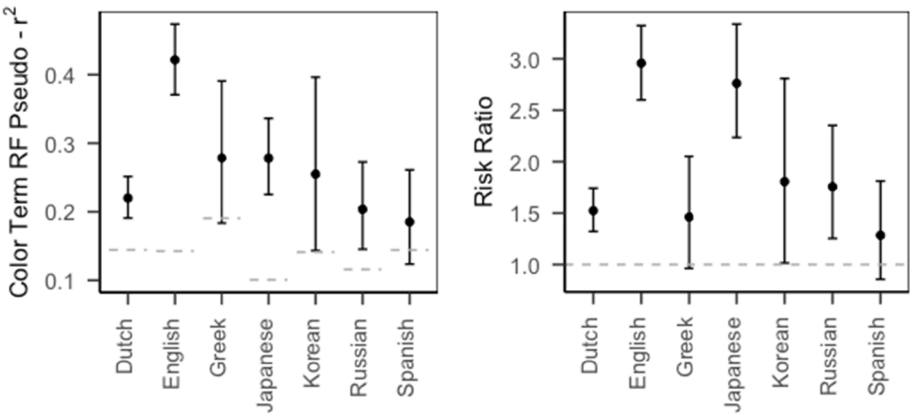

Fig. 1. a (left). For each language (x-axis), the pseudo-R^2 for the Color Term RF (black dot) and the pseudo-R^2 that would be expected under the null hypothesis (dotted line). Error bars are 95% confidence intervals. There is statistically significant evidence for the Color Term RF if the confidence interval does not cross the dotted line. Figure b (right). For each language (x-axis), the effect size of the Color Term RF, in units of risk ratio (black dot). Error bars are 95% confidence intervals. The dotted line at $RR = 1$ indicates "no effect"; there is statistically significant evidence for the Color Term RF in the language if the CI does not cross the dotted line.

Table 2
Descriptive and inferential statistics for the Color Term RF.

Language	Pseudo-R^2 (expected)	Pseudo-R^2 (observed)	Pseudo-R^2 (95% CI)	Pseudo-R^2 (p value)	Risk Ratio	Risk Ratio (95% CI)
Dutch	0.14	0.22	[0.19, 0.25]	<0.001	1.52	[1.32, 1.74]
English	0.14	0.42	[0.37, 0.47]	<0.001	2.96	[2.60, 3.32]
Greek	0.19	0.28	[0.18, 0.39]	=0.061	1.46	[0.96, 2.05]
Japanese	0.10	0.28	[0.23, 0.34]	<0.001	2.76	[2.24, 3.34]
Korean	0.14	0.26	[0.14, 0.40]	=0.026	1.81	[1.02, 2.81]
Russian	0.12	0.20	[0.15, 0.27]	<0.001	1.76	[1.25, 2.35]
Spanish	0.14	0.19	[0.12, 0.26]	=0.177	1.28	[0.86, 1.81]

is stronger in the second language.

To obtain a measure of effect size that allows accurate comparison between languages, we compute the Risk Ratio for the Color Term RF, dividing pseudo-R^2 by the expected value of pseudo-R^2 under the null hypothesis, $RR = R^2/E[R^2|H_0]$. The interpretation of risk ratio is quite intuitive; for example, a risk ratio $RR = 3$ for the Color Term RF means that on average "W" is three times more likely than other letters to be white, "Y" is three times more likely than other letters to be yellow, and so on. Furthermore, the comparison between two risk ratios is also intuitive: for example, the Color Term RF is twice as strong in a language with $RR = 3$ than in a language with $RR = 1.5$. Finally, the sampling distribution of log risk ratio is asymptotically normal (Agresti, 2003; Katz et al., 1978), and thus post-hoc pairwise comparisons (a test of whether the risk ratio in one language is significantly higher than in another) can be computed using z-tests of the difference between each pair of risk ratios (Altman & Bland, 2003).

Fig. 1b depicts the risk ratio for the Color Term RF in each language, along with the 95% confidence interval. Table 2 (right side) lists the effect size (in units of risk ratio) and confidence intervals for each language. Qualitatively, the effect size of the Color Term RF seems particularly high in English and Japanese. Consistent with this observation, post-hoc comparisons of the log-transformed risk ratio (with Bonferroni correction for the 21 possible comparisons) indicate that the effect size of the Color Term RF was significantly larger in English than in Dutch ($z = 7.04$, $correctedp < 0.001$), Spanish ($z = 4.15$, $correctedp < 0.001$), and Greek ($z = 3.47$, $correctedp = 0.011$), trending larger in English than in Russian ($z = 3.02$, $correctedp = 0.053$), significantly larger in Japanese than in Dutch ($z = 4.79$, $correctedp < 0.001$) and Spanish ($z = 3.54$, $correctedp = 0.009$), and trending larger in Japanese than in Greek ($z = 2.91$, $correctedp = 0.075$). No other post-hoc pairwise comparison was significant (all other $correctedp > 0.2$).

3.3. Discussion

In sum, we replicated Simner et al.'s (2005) original result in English, and also found evidence of a Color Term RF in four other languages (Dutch, Japanese, Korean, and Russian) plus a trending Color Term RF in Greek. However, we find no evidence of a Color Term RF in Spanish. In addition, the effect size of the Color Term RF was significantly larger in English and Japanese than in several other languages. Thus, the Color Term RF is not language-independent.

What is the implication of finding differences between languages in the effect size of the Color Term RF? The obvious answer is that single-language studies of Regulatory Factors must be interpreted with caution: we should not assume that an RF is language-independent (or language-dependent) until it has been tested in more than one language. More speculatively, one important

possibility is that some property of language influences the strength of the Color Term RF; in other words, "language matters". For example, it is possible that color terms are not a good "discriminating feature" (Asano & Yokosawa, 2013) in some languages: in Spanish "A is for *azul*", but "A is *also* for *amarillo*", so the Color Term RF does not yield uniquely-salient associations for each grapheme. Indeed, it is notable that the two languages in our dataset with the smallest number of unique initial graphemes for basic color terms are Greek and Spanish, and these are the two languages in our dataset with weak or absent Color Term RFs.

It is particularly interesting that the Color Term RF is quite strong in Japanese, despite the fact that Japanese adults typically write color terms using Kanji logographs (for traditional color terms) or Katakana (for foreign loanword color terms). However, at the age Japanese children acquire color terms (age 3–5; Saji et al., 2020), they still read and write using Hiragana, and begin to learn other Japanese scripts at a later age (Amano, 1986). Thus, our result could be explained by Japanese synesthetes acquiring these associations around age 5 or 6, the age at which synesthetic associations first start to develop (Simner et al., 2009) and also the age at which Japanese children use Hiragana spellings for color terms. This would also be consistent with Root et al.'s (2019) proposal that in synesthetes (but not non-synesthetes) the grapheme-color associations present in childhood are 'locked in' during their developmental trajectory. We therefore predict that if Japanese non-synesthete adults are given a forced-choice task to associate Hiragana graphemes to colors, they will show a weak or absent Color Term RF, whereas if Japanese non-synesthete children are given the same task, they will show a much stronger Color Term RF (similar in size to that of synesthetes). More generally, we predict that RFs based on properties that are learned or acquired early in development are particularly likely to be strong in synesthetes. In this way, differences between RF strengths in synesthetes and controls could offer insight into the developmental trajectory of letter representations.

That "language matters" (linguistic properties influence RF strength) is one possible (and attractive) interpretation of our result, but there are also several potential confounds. First, in some languages, a stronger RF might bias graphemes to be associated with colors that are incongruent with the Color Term RF. For example, we have previously shown that the first grapheme of the alphabet is red in many languages (Root et al., 2018); in our dataset, 13/21 (62%) of Spanish synesthetes associate "A" with red. If the 8 synesthetes who did *not* associate "A" with red all associated "A" with yellow (*amarillo*) or blue (*azul*), this would suggest that the "First Grapheme" RF is obscuring the effect of the Color Term RF. However, only 2/8 (25%) of these subjects associated "A" with yellow or blue, fewer than would be expected by chance (38%), though not significantly so (binomial test, $p = 0.072$). Therefore, we find it unlikely that the interference from other RFs is responsible for the null effect in Spanish. In Supplemental Text S3, we perform an additional *post hoc* exploratory analysis looking for evidence of RF interference, and do not find any evidence that our results are due to interference from incongruent RF predictions.

Another potential source of interference is from RFs in *a second language*, particularly English. Many – if not most – synesthetes in our study are proficient in English. Might English associations influence the colors experienced in our synesthetes' native languages? For example, Yellow is *Geel* in Dutch (predicting "G" is yellow), but it is not unreasonable to imagine that familiarity with English color terms could also influence Dutch synesthetes to associate "Y" with yellow (via the English "yellow"). In Supplemental Text S2, we explicitly test this hypothesis, but do not find any evidence that synesthetic associations in non-English languages are influenced by English color terms. This does not entirely rule out a potential effect of English second language proficiency, but we believe this is unlikely. First, the Color Term RF is very strong, and color terms are taught early in language instruction, so it is *a priori* the RF we would most expect to transfer from a second language. Second, in the countries in our study, English second language instruction typically begins around age 10, after many synesthetic associations have already been "locked in" (Simner & Bain, 2013).

A less interesting possibility is that our test is underpowered, and our trend in Greek and null effect in Spanish are Type II errors. Indeed, if the true effect size is equal to our observed effect size ($RR = 1.46$ in Greek and $RR = 1.28$ in Spanish), then we had only 46% power in Greek and 25% power in Spanish. However, even if our results are Type II errors, we have clear evidence that the Color Term RF is significantly *weaker* in Greek and Spanish. For both Greek and Spanish, we had >99.9% power to detect an effect as large as was observed in English, but failed to do so. Furthermore, the effect size in English was significantly larger than in Dutch, Greek and Spanish, and the effect size in Japanese was significantly larger than in Dutch and Spanish. These results cannot be explained by an underpowered sample, so statistical power is not a sufficient explanation for the language-dependence of the Color Term RF in our data.

Finally, another potential confound is that RFs *in general* could be stronger in some languages than others: it could be that we do not need to explain why the Color Term RF is stronger in English and weaker in Spanish; instead, we need to explain why RFs *in general* are strongest in English and weakest in Spanish. If overall RF effect size is responsible for our observed result, then there should be no "crossover interaction" between the Color Term RF and another RF. However, in the next section, we show that this is not the case: for a different RF, there is a strong effect in Spanish, but no effect in Korean.

4. Regulatory Factor 2: *Index Route RF* – "D is for dog and dogs are brown"

The influence of color terms on synesthetic associations is perhaps obvious, but synesthetic associations can also be influenced by semantic associations more indirectly. One notable example is the link between a grapheme, a word that begins with that grapheme (called "Index Words"; Mankin & Simner, 2017), and the prototypical color for the object (or concept) described by that word. For example, several researchers proposed that synesthetes associate the letter "A" with red because "A is for apple and apples are red" (Hancock, 2013; Mankin & Simner, 2017; Spector & Maurer, 2008; but see Root et al., 2018). Mankin and Simner (2017) were the first to quantify the effect of Index Words on synesthetic associations: they used data from non-synesthetes to generate predictions that were then tested on synesthetes. Non-synesthete controls generated words that begin with each letter, and a separate set of non-synesthete controls chose the prototypical color associated with the top three words for each letter. These results were combined to yield the colors that would be expected if the Index Route were driving synesthetic color, and these predictions were then compared to the actual

associations of synesthetes. They find that many more synesthetic associations are consistent with these predictions than would be expected by chance. Here, we test whether the results of Mankin and Simner (2017) can be replicated in all seven languages in our dataset.

4.1. Methods

To replicate Mankin and Simner (2017) in non-English languages, it was first necessary to collect letter-word and word-color associations from non-synesthetic control subjects in each of the languages in our dataset. It is necessary to generate Index Route predictions using data from non-synesthetes, rather than from synesthetes, as synesthetes' word choices might plausibly be influenced by the synesthetic colors they experience. The predictions derived from non-synesthetes are then compared to synesthetes' actual grapheme-color associations to determine the strength of the Index Route RF. We created seven translated versions for each of two experiments using the Qualtrics survey software (Qualtrics, 2013): one experiment to generate "index words" (letter to word associations), and a second experiment to determine the prototypical color of the most common index words (word to color associations).

For the first experiment, subjects (70 Dutch, 100 English, 111 Greek, 33 Japanese, 27 Korean, 112 Russian, and 45 Spanish subjects) provided letter-word associations for each grapheme in their native language. On each trial, a grapheme was presented, and the subject was prompted to type the first five words that "came to mind" that began with the grapheme. At the end of the experiment, each subject was screened for grapheme-color synesthesia, and any potential synesthete was excluded from the analysis. Synesthetes were excluded *post hoc* because we did not want to bias non-synesthetes responses by screening for synesthesia before the experiment. Our final dataset contained 53 Dutch, 65 English, 57 Greek, 26 Japanese, 27 Korean, 85 Russian, and 43 Spanish non-synesthetes. From this data, we chose the three most frequently chosen words for each grapheme in each language as "index words" (the same criterion as Mankin & Simner, 2017) for those graphemes.

For the second experiment, a different group of subjects (90 Dutch, 47 English, 108 Greek, 21 Japanese, 66 Korean, 57 Russian, and 37 Spanish subjects) provided word-color associations for each of their native language index words (i.e., three words per grapheme) from the previous experiment. On each trial, an index word was presented, and the subject was prompted to choose the Berlin-Kay color category that was the "best" color for that word. At the end of the experiment, each subject was screened for grapheme-color synesthesia, and any potential synesthete was excluded from the analysis (since synesthetes' choices might plausibly be influenced by their own synesthetic colors). We also excluded subjects that did not seem to take the task seriously, following the general method of Mankin and Simner (2017): we excluded any subject who chose the same color repeatedly (>3 standard deviations above the average number of times the color was chosen), or who chose the wrong color for color words (e.g., "Y is for yellow and yellow is black" is a clearly-incorrect response). Our final dataset contained 57 Dutch, 39 English, 68 Greek, 15 Japanese, 46 Korean, 40 Russian, and 29 Spanish subjects.

4.2. Results

Index word predictions were generated separately for each language (because, e.g., English speakers might consider apples to be red, whereas Dutch speakers consider them to be green). For each combination of index word and Berlin-Kay color, we multiplied the probability that the word was generated (in the first experiment) by the probability that the color was chosen for that word (in the second experiment). We excluded index words that were also Berlin-Kay color words (since we did not want to confound these results with those of the Color Term RF). We then took the sum of these probabilities for each combination of grapheme and Berlin-Kay color (i.e., collapsing across index words). Formally, for each grapheme g, color c, and index word w for that grapheme, we computed:

$$P(c|g) = \sum_{w \in RF_g} P(w)P(c|w)$$

From these probabilities, we used the criterion of Mankin and Simner (2017) to operationalize the Index Route RF: synesthetes should associate each grapheme with its *two most probable*[4] Index Route colors. This allowed us to use the same dataset (synesthetes' grapheme-color associations in each of the seven languages) and same analysis (binomial test on the observed *vs*. expected proportion of matches) as was used for the Color Term RF. We counted the number of trials that matched the predictions of the Index Route RF ($matches_{observed}$), the number of possible matches ($matches_{possible}$), and the expected proportion of matches under the null hypothesis ($E[R^2|H_0]$), and ran a binomial test of the hypothesis that $matches_{observed}$ successes out of a total of $matches_{possible}$ trials is consistent with the expected proportion of matches $E[R^2|H_0]$.

The proportion of associations consistent with the Index Route RF varied from 19% (Korean) to 32% (English). Fig. 2a depicts the expected pseudo-R^2, observed pseudo-R^2, and confidence intervals for each language; Table 3 (left side) lists these values along with p-values for the binomial test. The proportion of associations consistent with the Index Route RF was higher than would be predicted by chance in Dutch, English, Greek, Japanese, Russian, and Spanish, but was no higher than chance ($p > 0.999$) in Korean.

As with the Color Term RF, we quantified the effect size of the Index Route RF in units of risk ratio. Fig. 2b depicts the risk ratio for

[4] Although Mankin and Simner (2017) do not explicitly provide their rationale for using the two most probable colors rather than just the most probable color, one compelling reason to do so is that there are many graphemes for which there are two equally salient associations. For example, in our dataset, English "K" is predicted to be brown (via "kangaroo") 37% of the time and red (via "kite") 35% of the time.

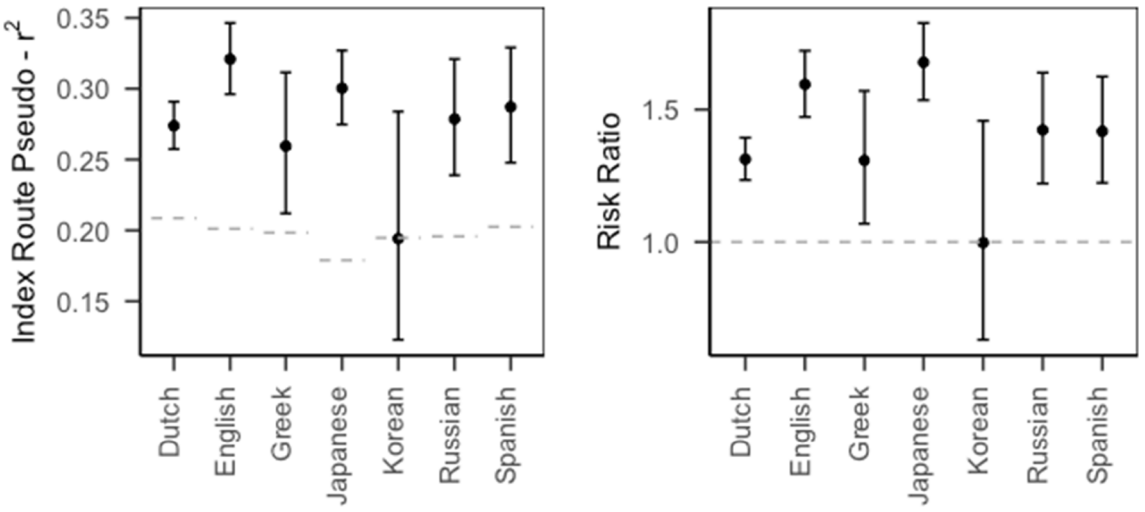

Fig. 2. a (left). For each language (x-axis), the pseudo-R^2 for the Index Route RF (black dot) and the pseudo-R^2 that would be expected under the null hypothesis (dotted line). Error bars are 95% confidence intervals. There is statistically significant evidence for the Index Route RF if the confidence interval does not cross the dotted line. Figure b (right). For each language (x-axis), the effect size of the Index Route RF, in units of risk ratio (black dot). Error bars are 95% confidence intervals. The dotted line at $RR = 1$ indicates "no effect"; there is statistically significant evidence for the Index Route RF in the language if the CI does not cross the dotted line.

Table 3
Descriptive and inferential statistics for the Index Route RF.

Language	Pseudo-R^2 (expected)	Pseudo-R^2 (observed)	Pseudo-R^2 (95% CI)	Pseudo-R^2 (p value)	Risk Ratio	Risk Ratio (95% CI)
Dutch	0.21	0.27	[0.26, 0.29]	<0.001	1.31	[1.23, 1.39]
English	0.20	0.32	[0.30, 0.35]	<0.001	1.60	[1.47, 1.72]
Greek	0.20	0.26	[0.21, 0.31]	=0.009	1.31	[1.07, 1.57]
Japanese	0.18	0.30	[0.28, 0.33]	<0.001	1.68	[1.54, 1.83]
Korean	0.20	0.19	[0.12, 0.28]	>0.999	0.997	[0.63, 1.46]
Russian	0.20	0.28	[0.24, 0.32]	<0.001	1.42	[1.22, 1.64]
Spanish	0.20	0.29	[0.25, 0.33]	<0.001	1.42	[1.22, 1.62]

the Index Route RF in each language, along with the 95% confidence interval. Table 3 (right side) lists the effect size (in units of risk ratio) and confidence intervals for each language. In addition, two post-hoc comparisons of the log-transformed risk ratio (with Bonferroni correction for the 21 possible comparisons) are significant: the effect size in English was significantly larger than in Dutch ($z = 3.85, corrected p = 0.003$), and the effect size in Japanese was significantly larger than in Dutch ($z = 4.54, corrected p < 0.001$). No other post-hoc pairwise comparison was significant (all other $corrected p > 0.2$).

Logically, there appears to be an interaction between RF and language: for example, the Color Term RF is significant in Korean but not Spanish; the Index Route RF is significant in Spanish but not Korean. We can test this formally using the tools of meta-analysis: we can interpret each risk ratio measurement as a separate "study" of the strength of RFs, and use the *rma* function in the *metafor* package for R (Viechtbauer, 2010) to run a fixed-effects meta-regression (Hedges & Vevea, 1998; van Houwelingen et al., 2002). By comparing the full model to a restricted model with interaction terms removed, we can test whether the heterogeneity in effect size across language can be attributed to the interaction between RF and language. Consistent with this hypothesis, the model that included all $RF \times Language$ interaction terms explained significantly more variance than the model that included just main effects of *RF* and *Language* (Likelihood Ratio Test, $\chi^2(6) = 28.58, p < 0.0001$). In sum, we find strong evidence for an interaction between RF and language, such that the relative strength of RFs is different in different languages.

4.3. Discussion

In sum, we replicated Mankin and Simner's (2017) original result in English[5], and also found evidence of an Index Route RF in five

[5] Although we replicate Mankin and Simner (2017), our Pseudo-R^2 is smaller (their reported Pseudo-R^2 is 0.40, which is outside our 95% CI for English, [0.30,0.35]). However, they measure the number of matches between the two most common index route colors for each grapheme and the *two most common* synesthetic colors for each grapheme, whereas we measure the number of matches between the index route and *all* synesthetic colors for each grapheme. It is thus unsurprising that our Pseudo-R^2 is different, and our results should not be interpreted as evidence for a weaker Index Route in English than was originally reported by Mankin and Simner (2017).

other languages (Dutch, Greek, Japanese, Spanish, and Russian). However, we find no evidence of an Index Route RF in Korean. Furthermore, the effect size of the Index Route RF in English and Japanese was significantly stronger than in Dutch. As with our Color Term RF results, the fact that we find differences between language in the effect size of the Index Route RF suggests that single-language studies of RFs must be interpreted with caution, and cannot be used to make universalist arguments about "how synesthetic associations come to be". Critically, using a meta-analytic comparison of the Index Route and Color Term RF results, we found an RF-by-language interaction in the observed effect size. This result eliminates one confound in our results for the Color Term RF: since we found a significant RF-by-language interaction, it cannot be the case that RFs *in general* are weaker in certain languages. Instead, a *particular* RF is stronger in a *particular* language.

Why might the Index Route RF be absent (or too weak to detect) in Korean? One possibility is that there is a cultural effect: in English-speaking countries schoolchildren often learn their letters using a relatively-consistent set of acrophonics ("A is for apple", "B is for boy", etc.), whereas that this may not be the case for all languages in our dataset. If a language uses consistent acrophonic words for letters, then between-subject agreement of the non-synesthete controls will be high on the word generation task. Indeed, the average between-subject agreement amongst non-synesthetes in our word generation task was highest in Japanese (36%) and English (30%), the two languages in our data for which the Index Route RF was strongest. This *post hoc* result suggests that the Index Route RF might be stronger in languages with consistent letter-to-word associations in the developmental environment. This hypothesis makes a very clear prediction about a language not in our dataset: Thai schoolchildren learn their writing system using official, standardized acrophonic words, most of which are highly-imageable objects ("chicken", "elephant", "teeth", etc.) with prototypical colors, between-subject agreement for the Thai word-generation task should be close to 100%. If the Index Route is influenced by the consistency of letter-to-word associations in the developmental environment, then we predict that the Index Route should be particularly strong in Thai synesthetes.

Japanese non-synesthete controls not only tend to agree on their word choices in the word generation task, they also chose highly-imageable words for which there was more agreement about the prototypical color (e.g., "banana" is highly-imageable and has a prototypical color, whereas "democracy" is not imageable and does not have a prototypical color). This suggests that the large effect size of the Index Route RF in Japanese might be due to the degree to which these associations are "discriminating features" of Hiragana graphemes (Asano & Yokosawa, 2013). In this framework, graphemes are influenced by whichever RF makes predictions that maximize discriminability between graphemes, and one prerequisite for such discriminability is that the RF make strong predictions that a grapheme will be a particular color. This hypothesis is tested *within*-language (in English synesthetes) on page 31 of Mankin and Simner (2017; also see their Fig. 7), who find that in English, the graphemes that are likeliest to be influenced by the Index Route are those for which the RF predicts a single dominant color. We suggest that this may also be true *between*-language: that languages in which the Index Route RF makes strong color predictions for every grapheme are languages in which the effect size of the Index Route RF will be strongest.

As with the Color Term RF, one potential confound is that a stronger RF in Korean overpowers any influence of the Index Route. For example, Hangul is a "featural" alphabet that is heavily structured around sound (Haarmann, 1993), and sound similarity exerts a very strong influence on Korean grapheme-color associations (Kang et al., 2017), and the predictions of this RF could be particularly incongruent with those of the Index Route. We also noticed that some statistical features of the Index Route data itself seemed to be related to RF strength. For example, Korean synesthetes experience 16% of graphemes *in general* as yellow, but only 10% of Korean Index Route predictions are of yellow. If synesthetes are biased to experience yellow associations *in general* (which is certainly true in English and Dutch; Rouw & Root, 2019), then this might reduce the effect size of all non-yellow index route words in Korean.

As with the Color Term RF, it is also possible that our test is underpowered. Effect size estimates for the Index Route RF are *in general* likely to be underestimates that can be thought of as a "floor" on the true effect size, because measuring index words at the level of the adult population will miss idiosyncratic but salient individual associations (e.g., a synesthete who grew up in a household of Oregon Ducks football fans, and learns that "O" is yellow), and may also miss associations that are common in children but rarer in adults. However, we do not have any *a priori* reason to believe that our method (measuring adult associations in non-synesthetes) should yield *different* levels of underestimation for each language. The large difference in sample sizes between language may be more concerning; indeed, the small sample size ($n = 13$) of Korean synesthetes yields quite large confidence intervals on our estimates of effect size (c.f. Fig. 2, and also Figs. 1 and 3). However, the point estimate for the risk ratio in Korean is actually *below* the null hypothesis risk ratio of 1.0, so we find it unlikely that a larger Korean sample would change the result. Indeed, even at the current sample size, we have 87% power in Korean to detect an effect that is as large as the observed effect size in Japanese, and failed to do so. Finally, the significant difference in effect size between language and the *RF* × *Language* interaction cannot be explained by an underpowered sample; in other words, insufficient sample size is not a sufficient explanation for the language-dependence of the Index Route RF in our data.

Thus far, we have shown that two different RFs seem language-dependent: not only is there a significant difference in effect size between languages, there is also a significant *interaction* between RF and language. It is not simply the case that RFs are weaker in Spanish or that the Index Route RF is stronger than the Color Term RF; instead, certain RFs are stronger in certain languages. Does this then mean that all RFs are language-dependent? Or, might some RFs exert their influence equally in all languages? The third RF we examine is that which we expect *a priori* to have the best chance of being truly language independent.

5. Regulatory Factor 3: *Basic Shape RF* – pre-linguistic associations

Most RFs are linguistic in origin (color terms, the Index Route, letter frequency, etc.), but some RFs are related to the visual shape of letters rather than their meaning. For example, Spector and Maurer (2008) found that pre-literate children (who did not yet know how to read, and could not match the names of letters to their visual shape) nevertheless associated the shapes "O" and "I" with white, and

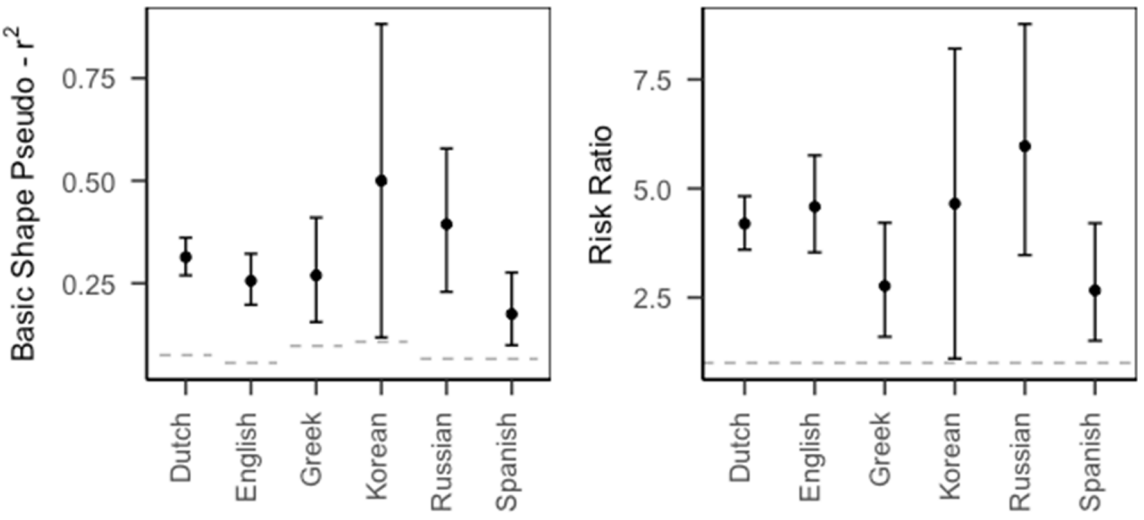

Fig. 3. a (left). For each language (x-axis), the pseudo-R^2 for the Basic Shape RF (black dot) and the pseudo-R^2 that would be expected under the null hypothesis (dotted line). Error bars are 95% confidence intervals. There is statistically significant evidence for the Basic Shape RF if the confidence interval does not cross the dotted line. Figure b (right). For each language (x-axis), the effect size of the Basic Shape RF, in units of risk ratio (black dot). Error bars are 95% confidence intervals. The dotted line at $RR = 1$ indicates "no effect"; there is statistically significant evidence for the Basic Shape RF in the language if the CI does not cross the dotted line.

the shapes "X" and "Z" with black, which they attributed to the visual properties of the letters ("simple" *vs.* "complex", "ameboid" *vs.* "jagged"; Spector & Maurer, 2011). In contrast, pre-literate children did *not* associate "G" with green (since this effect requires knowing that "G" is the initial letter of the word "green"). Spector and Maurer suggest that this prelinguistic effect influences synesthetes' associations for "O", "I", "X", and "Z", although they do not measure these associations in synesthetes themselves. Since this effect is prelinguistic, we predict it to be language-independent in synesthetes: this "Basic Shape" RF should influence synesthetes' grapheme-color associations in any language that contains these shapes. As far as we know, this is the first time that a prelinguistic influence on synesthetic associations has been compared across languages.

5.1. Methods

We chose to operationalize the Basic Shape RF strictly: we restricted our analysis to shapes that exactly matched the shapes of "X", "Z", "O", and "I". These four letters are all in the English, Dutch, Spanish, and Greek alphabets, and "O" and "X" are in the Russian alphabet. In addition, although Korean Hangul script does not descend from the Latin alphabet, it contains two characters that are visually identical to "O" and "I": "O" and "|". However, we excluded letters such as Greek Θ and Russian Ж, because their shapes are not identical to "O" and "X". In addition, we excluded Japanese from the analysis because there are no Hiragana characters with these shapes. Our hypothesis was that the shapes "O" and "I" would be white more often than chance, and that the shapes "X" and "Z" would be black more often than chance.

5.2. Results

We used the same analysis as with the Color Term and Index Route RF: we counted the number of trials that matched the predictions of the Basic Shape RF ($matches_{observed}$), the number of possible matches ($matches_{possible}$), and the expected proportion of matches under the null hypothesis ($E[R^2|H_0]$), and ran a binomial test of the hypothesis that $matches_{observed}$ successes out of a total of $matches_{possible}$ trials is consistent with the expected proportion of matches $E[R^2|H_0]$.

The proportion of associations consistent with the Basic Shape RF varied from 18% (Spanish) to 50% (Korean). Fig. 3a depicts the expected and observed pseudo-R^2 and their confidence intervals; Table 4 (left side) lists these values along with p-values for the binomial test. The proportion of associations consistent with the Basic Shape RF was higher than would be predicted by chance in all six languages tested. As with the Color Term and Index Route RFs, we quantified the effect size of the Basic Shape RF in units of risk ratio. Fig. 3b depicts the risk ratio and 95% confidence interval for the Basic Shape RF in each language, and Table 4 (right side) lists their numeric values. Unlike the Color Term and Index Route RFs, no post-hoc comparisons of the log-transformed risk ratio (Bonferroni-corrected for 15 comparisons) were significant (all $correctedp > 0.2$).

5.3. Discussion

In sum, we found that Spector and Maurer's (2008) result in pre-linguistic children was consistent with synesthetic associations in

all examined languages; English, Dutch, Greek, Korean, Spanish, and Russian. Crucially, Spector and Maurer's result that these associations exist in *prelinguistic* infants (who do not know any graphemes) predicts that the Basic Shape RF should not be language-dependent. Consistent with this prediction, we found that the Basic Shape RF does *not* significantly differ in effect size between language.

In this analysis, we chose to strictly operationalize the Basic Shape RF as "graphemes with the exact shapes O, I, X, or Z", and therefore could not include Japanese in our analysis. Although Spector and Maurer suggested that it is the "jagged"-ness (*vs.* "amoeboid"-ness) or the complexity (e.g., the number of intersections) of letter shapes that induces them to be black or white, they do not operationalize these descriptors in such a way that predictions about Japanese graphemes can be made. We do find it notable that the Hiragana graphemes most frequently associated with white (し/つ/ひ) seem to be subjectively "simple", and the Hiragana graphemes most frequently associated with black (ん/お/を) seem to be subjectively more "complex", but it is quite possible that this is confirmation bias on our part. Furthermore, it is unclear why "Z" is black, but not "N", considering that these two letters are visually identical after a 90-degree rotation. Future research should determine what specific, measurable features of these shapes cause them to be associated with specific colors. The Basic Shape RF could then be operationalized in such a way that it could be tested in more languages with non-Latin scripts.

By combining Spector and Maurer's (2008) developmental result with data from synesthetes across languages, we found that some grapheme-color associations are influenced by "universal", naturally biased associations. As far as we know, this is the first time that an influence on grapheme-color associations has been shown to be prelinguistic. More generally, our results provide the exciting suggestion that influences on the development of letter representations can be "traced" by comparing grapheme-color associations across different ages and different languages.

6. General discussion

We characterized the degree to which regulatory factors – factors that influence grapheme-color associations in synesthesia – are similar or different in synesthetes with different native languages. Overall, we found clear evidence that RFs can be language-specific: we found statistically significant differences in the effect size of RFs between languages, and in two cases, we found no evidence for a particular RF in a particular language. Importantly, the effect size was dependent on the *interaction* between RF and language. It is not simply the case that some RFs are stronger in general, nor is it simply the case that all RFs are stronger in a certain language: for example, we found evidence for the Color Term RF in Korean but not Spanish, and evidence for the Index Route RF in Spanish but not Korean. We also found evidence that RFs can be language-independent: in all languages tested, graphemes with the shapes O/I and X/Z were likelier to be white and black, respectively, and the effect size of this RF was not significantly different in different languages.

Our results suggest that monolingual studies of RFs should be interpreted with caution. When an RF is found to exert strong influence on grapheme-color associations in a single language, we should not assume that this reflects a universal truth of synesthetic associations: the effect might instead be explained by properties of the *particular* language or properties of the environment (e.g., educational strategies in a particular country). Our results also have implications for a "file drawer" effect, in which an RF tested in a single language yields a null result and is not published or publicized. It is possible that an interesting RF is "sitting in a file drawer" because *it was tested in the wrong language*. Indeed, it is noteworthy that the effect sizes in our study were particularly strong in English. It is of course possible that RFs are truly stronger in English than in other languages, *in general*. It is also of course possible that the observed trend for stronger effect sizes in English is a coincidence. However, an additional possibility is that there is a "file drawer" effect: most published studies of synesthesia use English-speaking synesthetes, so published RFs will be those that are particularly strong in English (strong enough to be statistically significant). To overcome this potential source of bias, null results of RF experiments (including confidence intervals on effect size) should be published, or at least registered in a repository. One additional possibility which we find particularly intriguing is that there is a bias in synesthesia studies because English-speaking *researchers*, when brainstorming potential RFs to test, are likelier to think of RFs that are intuitive to English speakers (and thus likely to be particularly strong RFs in English). To overcome this potential source of bias, the synesthesia research community should seek to foster interest and growth in synesthesia research at universities in non-English-speaking countries.

More speculatively, we wonder if cross-linguistic differences in synesthesia could extend beyond just color associations: could there be cross-linguistic *perceptual* differences in the experience of synesthesia? For example, might test–retest consistency be lower or higher in different languages? In the analyses presented here, we use the same test–retest consistency (≤ 135) in each language because our aim was to apply an identical analysis pipeline across different languages and RFs, but it is entirely possible that the ideal threshold to separate synesthetes from non-synesthetes is different in different languages. This is particularly relevant when considering the effect

Table 4
Descriptive and inferential statistics for the Basic Shape RF.

Language	Pseudo-R^2 (expected)	Pseudo-R^2 (observed)	Pseudo-R^2 (95% CI)	Pseudo-R^2 (p value)	Risk Ratio	Risk Ratio (95% CI)
Dutch	0.07	0.31	[0.27, 0.36]	<0.001	4.19	[3.60, 4.82]
English	0.06	0.26	[0.20, 0.32]	<0.001	4.58	[3.53, 5.76]
Greek	0.10	0.27	[0.16, 0.41]	<0.001	2.76	[1.60, 4.21]
Korean	0.11	0.50	[0.12, 0.88]	=0.019	4.65	[1.10, 8.21]
Russian	0.07	0.39	[0.23, 0.58]	<0.001	5.97	[3.47, 8.77]
Spanish	0.07	0.18	[0.10, 0.28]	<0.001	2.66	[1.51, 4.20]

of the threshold on our sample: using a threshold that is too high may lead us to include non-synesthetes who have some degree of consistency; using a threshold that is too low may lead us to exclude true synesthetes. Future work can replicate the methods of Rothen et al. (2013) in synesthetes with different native languages to test whether average test–retest consistency is language-specific. In addition, the relationship between the colors that synesthetes experience for *graphemes* and the colors that they experience for *words* is often complex (e.g., Blazej & Cohen-Goldberg, 2016; Mankin et al., 2016). Accounting for cross-language differences in the relationship between grapheme and word color would be a complex endeavor, but could yield many more insights than studying grapheme color alone. Furthermore, might the actual *conscious perception* of synesthetes be different in different languages? For example, could the ratio of associators (synesthetes who see the color "in their mind's eye") to projectors (synesthetes who actually see the color on top of the letter) be higher in languages with stronger semantic (e.g., Index Route) *vs* visual (e.g., Basic Shape) RFs? Some visual illusions are famously stronger or weaker in certain cultures (Segall et al., 1966), so such differences are not beyond the realm of possibility. If there are indeed cross-language differences in the perceptual quality of synesthesia, this would offer a powerful example of the capacity for language to shape perception (linguistic relativity; for a review, see Lupyan et al., 2020). More generally, data from large international samples could help disentangle genetic and environmental contributions to the synesthetic experience: cross-language differences in synesthesia illuminate the extent to which environment can influence conscious perceptual experience.

One important limitation of our study is that sample size is very limited in several languages. It is plausible that one or more of the null results we report here is a Type II error. However, this is not a problem for our conclusion, because our argument that language matters does not rely on showing that an RF exerts *truly no influence* (effect size of zero) in certain languages: we also show that RFs are *stronger or weaker* in certain languages. Even with small sample sizes for some languages, we have clearly shown that the effect sizes of RFs are significantly different across language. Confidence intervals on effect sizes are a more interesting and important measure of RFs than whether or not an RF is "significant", and the insights to be gleaned from future cross-language studies of RFs will likely come from comparing effect sizes across language rather than from null hypothesis significance testing in each language.

A more complex limitation of our study is that although we have modeled each RF separately, different RFs make predictions about the same underlying data. If RFs make congruent predictions about a grapheme, our effect sizes for those RFs might be overestimated; if RFs make incongruent predictions about a grapheme, our effect sizes for those RFs might be underestimated. For example, the grapheme in the first ordinal position tends to be red (Root et al., 2018), but the Color Term RF predicts that Spanish "A" should be blue (*azul*) or yellow (*amarillo*). In addition, Russian "O" is *particularly* likely to be white; this could be because the Basic Shape RF is particularly strong, or it could be because both the Basic Shape RF ("O-shapes are white") and the Index Route RF ("O" → "облако" (cloud) → white) are simultaneously biasing Russian "O" to be white. Here, we used some *post hoc* analyses to examine these possibilities (e.g., showing that Spanish synesthetes who do not experience a red "A" *also* do not experience a yellow or blue "A"). However, to completely eliminate this confound, it would be necessary to allow RFs to interact as multiple predictors in a single model. Furthermore, all RFs modeled in this study are *first order* RFs: they make predictions about which color is associated with grapheme. Other RFs are *second order*: they make predictions about which graphemes share similar colors. For example, one second-order RF that is predicted to be language-dependent (Asano & Yokosawa, 2013) is that similarly-pronounced graphemes are associated with similar colors only in languages that are orthographically transparent (i.e., has a phonemic writing system in which there is a high degree of consistency between a grapheme and its pronunciation). Consistent with this hypothesis, pronunciation does not influence synesthetic color in English (an opaque orthography; Watson et al., 2012), but does influence synesthetic color in Japanese Hiragana and Korean Hangul (both transparent orthographies; Asano & Yokosawa, 2013; Kang et al., 2017). Unfortunately, this "Pronunciation RF" (and other second-order RFs) cannot be operationalized in such a way that it can be analyzed using the technique used in the present study.

In the future, researchers should work to build a larger, high-powered, cross-language database of synesthetic subjects and should quantify the effect of all RFs within a single model. It is an open question how such a model should be specified. Should the model assume that RFs exert influence equally on all graphemes (*vs.* varying influence based on grapheme "discriminating features"; Asano & Yokosawa, 2013)? How do RFs interact when they make conflicting or congruent predictions? How do the *first-order* RFs (RFs that predict a single grapheme should be a particular color – such as those measured in this study) interact with *second-order* RFs? These decisions involve a large number of "experimenter degrees of freedom" (Simmons et al., 2011), so it is important that they be made *a priori* using pilot analysis of existing datasets, or that they be made on a subset of collected data and cross-validated using holdout data.

Our current results strongly suggest that such an undertaking is worthwhile: we find that although some Regulatory Factors may be universal, some differ between language. We suggest that cross-language differences in RF strength could be attributable to cross-language differences in properties of the language (such as orthographic depth), properties of the RF (such as the imageability of each language's index words) and properties of the environment (such as the use of acrophonic teaching in schools). In this way, cross-language differences in grapheme-color associations are not noise, or a covariate to be controlled for, but are instead a signal that – when understood – will yield insights into how language influences grapheme representation in the brain.

CRediT authorship contribution statement

Nicholas Root: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Michiko Asano:** Investigation, Data curation, Resources, Writing – review & editing. **Helena Melero:** Investigation, Data curation, Resources, Writing – review & editing. **Chai-Youn Kim:** Investigation, Data curation, Resources, Writing – review & editing. **Anton V. Sidoroff-Dorso:** Investigation, Data curation, Resources, Writing – review & editing. **Argiro Vatakis:** Investigation, Data curation, Resources, Writing – review & editing. **Kazuhiko Yokosawa:** Supervision, Writing – review & editing. **Vilayanur Ramachandran:** Investigation, Supervision, Writing – review & editing. **Romke Rouw:** Investigation, Data curation, Writing – original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank David Brang for help with American data collection and sharing, and Ji-won Siong and Youjung Lee for their help with Korean data collection and sharing, and Jelle Zuidema for helpful conversations about the methodology. This research was supported by the University of Amsterdam (Amsterdam Brain and Cognition Project Grant 06_PG19 awarded to authors NR and RR), the Centre of Interdisciplinary Research of Contemporary Childhood of Moscow State University of Psychology and Education (the *Synaesthesia: Children and Parents* project led by author AVSD), the Japan Society for the Promotion of Science (JSPS KAKENHI 19H01770 awarded to authors MA and KY), and the National Research Foundation of Korea (NRF-2020R1F1A1076336 awarded to author CYK).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2021.103192.

References

Agresti, A. (2003). *Categorical data analysis* (Vol. 482). John Wiley & Sons.
Altman, D. G., & Bland, J. M. (2003). Interaction revisited: The difference between two estimates. *Bmj, 326*(7382), 219.
Amano, K. (1986). *Learning of Kana in children*. Tokyo: Akiyama-shoten (in Japanese).
Asano, M., & Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes. *Consciousness and Cognition, 20*(4), 1816–1823.
Asano, M., & Yokosawa, K. (2013). Grapheme learning and grapheme-color synesthesia: Toward a comprehensive model of grapheme-color association. *Frontiers in Human Neuroscience, 7*, 757.
Asano, M., Nagai, J., & Yokosawa, K. (2017). Temporal consistency in grapheme-color synesthesia covaries with sensitivity to regulatory factors in grapheme-color associations. In International association of synaesthetes, artists, and scientists (IASAS)' first synaesthesia symposium. Los Angeles, USA (October 20, 2017).
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.
Владение иностранныМи языкаМи (2014, May 28). Levada Center. https://www.levada.ru/2014/05/28/vladenie-inostrannymi-yazykami/.
Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution*. Univ of California Press.
Blazej, L. J., & Cohen-Goldberg, A. M. (2016). Multicolored words: Uncovering the relationship between reading mechanisms and synesthesia. *Cortex, 75*, 160–179.
Brang, D., Hubbard, E. M., Coulson, S., Huang, M., & Ramachandran, V. S. (2010). Magnetoencephalography reveals early activation of V4 in grapheme-color synesthesia. *Neuroimage, 53*(1), 268–274.
Chromý, J., Borůvková, M., Malá, L., & Sudzinová, T. (2019). Long-term versus short-term consistency in the grapheme–colour synaesthesia: Grapheme–colour pairings can change in adulthood. *Attention, Perception, & Psychophysics, 81*(6), 1805–1812.
Coventry, K., Mitsakis, C., Davies, I., Jover, J. L., Androulaki, †., & Gómez-Pestaña, N. (2006). Basic colour terms in Modern Greek: Twelve terms including two blues. *Journal of Greek Linguistics, 7*(1), 3–47.
Davies, I., & Corbett, G. (1994). The basic color terms of Russian. *Linguistics, 32*(1), 65–90.
Day, S. A. (2004). Trends in synesthetically colored graphemes and phonemes—2004 revision. Retrieved from http://www.daysyn.com/Day2004Trends.pdf.
Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods, 159*(1), 139–145.
Hancock, P. (2013). Synesthesia, alphabet books, and fridge magnets. In J. Simner, & E. M. Hubbard (Eds.), *The oxford handbook of synesthesia* (pp. 83–102). Oxford University Press.
Haarmann, H. (1993). The emergence of the Korean script as a symbol of Korean identity. *Contributions to the Sociology of Language, 65*, 143-143.
Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486–504.
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.
Hupé, J. M., Bordier, C., & Dojat, M. (2012). The neural bases of grapheme–color synesthesia are not localized in real color-sensitive areas. *Cerebral Cortex, 22*(7), 1622–1633.
Kang, M. J., Kim, Y., Shin, J. Y., & Kim, C. Y. (2017). Graphemes sharing phonetic features tend to induce similar synesthetic colors. *Frontiers in Psychology, 8*, 337.
Katz, D., Baptista, J., Azen, S. P., & Pike, M. C. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics, 34*(3), 469. https://doi.org/10.2307/2530610.
Kim, C. Y., & Blake, R. (2013). Revisiting the perceptual reality of synesthetic color. In *Oxford handbook of synesthesia* (p. 283). Oxford University Press.
Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. https://doi.org/10.31234/osf.io/pztmc.
Mankin, J. L., Thompson, C., Branigan, H. P., & Simner, J. (2016). Processing compound words: Evidence from synaesthesia. *Cognition, 150*, 1–9.
Mankin, J. L., & Simner, J. (2017). A is for apple: The role of letter–word associations in the development of grapheme–colour synaesthesia. *Multisensory Research, 30*(3-5), 409–446.
Mankin, J. L. (2019). Deepening understanding of language through synaesthesia: A call to reform and expand. *Philosophical Transactions of the Royal Society B, 374*(1787), 20180350. https://doi.org/10.1098/rstb.2018.0350.
Melero, H., Ríos-Lago, M., Peña-Melián, A., & Álvarez-Linera, J. (2014). Achromatic synesthesias—A functional magnetic resonance imaging study. *Neuroimage, 98*, 416–424.
Melero, H., Peña-Melián, Á., & Rios-Lago, M. (2015). Colors, tastes, numbers? Synesthesia in a Spanish sample. *Revista de neurología, 60*(4), 145.
Mylonas, D., MacDonald, L., & Wuerger, S. (2010, January). Towards an online color naming model. In Color and imaging conference (Vol. 2010, No. 1, pp. 140–144). Society for Imaging Science and Technology.
Newell, F. N., & Mitchell, K. J. (2016). Multisensory integration and cross-modal learning in synaesthesia: A unifying model. *Neuropsychologia, 88*, 140–150.
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. *Nature Neuroscience, 5*(4), 371–375.

Palmeri, T. J., Blake, R., Marois, R., Flanery, M. A., & Whetsell, W. (2002). The perceptual reality of synesthetic colors. *Proceedings of the National Academy of Sciences, 99*(6), 4127–4131.
Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online color naming experiment in Russian using Munsell color samples. *Color Research & Application, 43*(3), 358–374.
Qualtrics, I. (2013). *Qualtrics*. UT, USA: Provo.
Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.
Regier, T., Kay, P., & Cook, R. S. (2005). Universal foci and varying boundaries in linguistic color categories. In Proceedings of the 27th meeting of the cognitive science society (pp. 1827–1832). Hillsdale (NJ): Cognitive Science Society.
Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods, 215*(1), 156–160.
Rich, A. N., Bradshaw, J. L., & Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical-colour associations. *Cognition, 98*(1), 53–84.
Root, N. B., Dobkins, K., Ramachandran, V. S., & Rouw, R. (2019). Echoes from the past: Synaesthetic colour associations reflect childhood gender stereotypes. *Philosophical Transactions of the Royal Society B, 374*(1787), 20180572. https://doi.org/10.1098/rstb.2018.0572.
Root, N. B., Rouw, R., Asano, M., Kim, C. Y., Melero, H., Yokosawa, K., & Ramachandran, V. S. (2018). Why is the synesthete's "A" red? Using a five-language dataset to disentangle the effects of shape, sound, semantics, and ordinality on inducer–concurrent relationships in grapheme-color synesthesia. *Cortex, 99*, 375–389.
Rouw, R., Case, L., Gosavi, R., & Ramachandran, V. (2014). Color associations for days and letters across different languages. *Frontiers in Psychology, 5*, 369.
Rouw, R., & Root, N. B. (2019). Distinct colours in the 'synaesthetic colour palette'. *Philosophical Transactions of the Royal Society B, 374*(1787), 20190028. https://doi.org/10.1098/rstb.2019.0028.
Saji, N., Imai, M., & Asano, M. (2020). Acquisition of the meaning of the word orange requires understanding of the meanings of red, pink, and purple: Constructing a lexicon as a connected system. *Cognitive Science, 44*(1). https://doi.org/10.1111/cogs.v44.110.1111/cogs.12813.
Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception* (p. 32). Indianapolis: Bobbs-Merrill.
Shin, E.-H., & Kim, C.-Y. (2014). Both "나" and "な" are yellow: Cross-linguistic investigation in search of the determinants of synesthetic color. *Neuropsychologia, 65*, 25–36.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.
Simner, J. (2007). Beyond perception: synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences, 11*(1), 23–29.
Simner, J., & Bain, A. E. (2013). A longitudinal study of grapheme-color synesthesia in childhood: 6/7 years to 10/11 years. *Frontiers in Human Neuroscience, 7*, 603.
Simner, J., Harrold, J., Creed, H., Monro, L., & Foulkes, L. (2009). Early detection of markers for synaesthesia in childhood populations. *Brain, 132*(1), 57–64.
Simner, J., Ipser, A., Smees, R., & Alvarez, J. (2017). Does synaesthesia age? Changes in the quality and consistency of synaesthetic associations. *Neuropsychologia, 106*, 407–416.
Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology, 22*(8), 1069–1085.
Spector, F., & Maurer, D. (2008). The colour of Os: Naturally biased associations between shape and colour. *Perception, 37*(6), 841–847.
Spector, F., & Maurer, D. (2011). The colors of the alphabet: Naturally-biased associations between shape and color. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 484–495.
van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine, 21*(4), 589–624.
van Leeuwen, T. M., Dingemanse, M., Todil, B., Agameya, A., & Majid, A. (2016). Nonrandom associations of graphemes with colors in Arabic. *Multisensory Research, 29*(1-3), 223–252.
Veall, M. R., & Zimmermann, K. F. (1992). Performance measures from prediction-realization tables. *Economics Letters, 39*(2), 129–134.
Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R2 measures for some common limited dependent variable models. *Journal of Economic surveys, 10*(3), 241–259.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.
Watson, M. R., Akins, K. A., & Enns, J. T. (2012). Second-order mappings in grapheme–color synesthesia. *Psychonomic Bulletin & Review, 19*(2), 211–217.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780–7785.