Running head: FREQUENCY-BASED SYNESTHETIC ASSOCIATIONS

Why Q is Purple:

Frequency-Based Synesthetic Associations between Letters and Colors

Laura M. Herman

Department of Psychology, Harvard University,
Pine Crest School,
Department of Psychology, Princeton University,

Jordan W. Suchow

Department of Psychology, Harvard University,
Institute of Cognitive and Brain Sciences, University of California, Berkeley

Abstract

People with grapheme–color synesthesia associate numbers and letters with colors. Each grapheme is typically associated with a specific color, but the cause of these pairings is unknown. A useful clue is that the pairings are consistent among people who use the same language, but differ across languages. Here, we show that synesthetes' exposure to written language guides the grapheme–color associations that they form. Specifically, we found that among English-, Spanish-, and German-speaking synesthetes, the hue associated with a letter is strongly correlated with the frequency of that letter in written language. Letters that are more frequent are associated with red, orange, and yellow; less frequent letters are associated with green, blue, and violet. Conversely, when asked to assign a color to each letter, non-synesthetes associated colors with letters differently, showing a weak correlation between letter frequency and hue and often assigning colors by semantic and orthographic association (e.g., Y/yellow or A/apple/red). This provides evidence that synesthetic associations are distinct from those of the general population. In light of previous work, these results suggest that synesthetic associations between graphemes and colors originate from exposure to text, fossilize in childhood, and remain stable throughout life.

Introduction

Synesthesia is a neurological condition in which the stimulation of one sense leads to the automatic experience of another sense. In grapheme–color synesthesia, for example, an individual's perception of each number and letter is associated with a color (Baron-Cohen, Wyke, & Binne, 1987). Synesthesia is heritable and is estimated to affect 2–3% of the population (Simner et al., 2006). Many synesthetes have an aptitude for the arts, a strong sense of creativity, and above-average memory skills (Baron-Cohen et al., 1987; Smilek, Dixon, Cudahy, & Merikle, 2002).

Synesthetes do not simply imagine or purposefully associate colors with letters. One line of evidence for this comes from comparisons of the neural activity of natural word–color synesthetes to so-called "forced synesthetes" — people who have been trained in the lab to recall a certain color whenever a paired letter or word is heard. Specifically, visual areas V4 and V8, used in color perception, are more active in natural synesthetes than in forced synesthetes (Nunn et al., 2002). This is consistent with models of synesthesia in which it arises from communication between cortical areas that are not strongly connected in non-synesthetes (Bargary, 2007).

In grapheme–color synesthesia, each grapheme is typically associated with a specific color, but the cause of these pairings is unknown. Many have considered whether synesthetic associations arise from childhood experience with toys and other artifacts (e.g. a kitchen magnet set where A is red), from semantic associations (e.g. A is for apple), or from other synesthetic family members (Barnett et al., 2008; Marks, 1975; Rich, Bradshaw, & Mattingley, 2005; Simner et al., 2005; Hancock, 2005; Witthoft & Winawer, 2005; Ward & Simner, 2003).

A useful clue for understanding the causes of grapheme–color pairings is that they are consistent among people who use the same language, but differ across languages. The "luminance" of synesthetic colors is known to depend on the frequency of letters and digits in everyday language, as does their saturation, demonstrating that synesthetic associations are affected by the prevalence of the grapheme (Beeli, Esslen, & Jancke, 2007). Here, we carry that idea a step farther by considering the relationship between a letter's prevalence in English, Spanish, and German and the hue of the color most commonly associated with it, finding frequency-based synesthetic associations that explain much of the observed pattern of grapheme–color pairings that is seen across the three languages.

Experiment 1

Methods

In Experiment 1, we measured the correlation between the frequency of a letter in English, Spanish, and German and the hue of the color associated with it by a majority of synesthetes who use the respective language.

Synesthetic Letter–Color Associations. Letter-color associations for the English language were derived from the *Common Letter-Colour Associations* chart, which aggregates color–letter associations found in other studies, highlighting cases where the majority of synesthetes have the same association (Barnett et al., 2008).

For the Spanish language, members of a deidentified pool of Spanish grapheme-color synesthetes at the University of Granada were asked to choose the color that most closely matches their association with each letter (*n*=14). Each participant submitted a document with each letter printed in the associated color. The hue of each reported color was measured. For each letter, we found the circular mean of the reported hue across all synesthetes.

Similarly, for the German language, we used data from an existing study to deduce the hues of the colors that German synesthetes (*n*=89) most typically associate with letters (Emrich, Schneider, & Zedler, 2004). This study reported the majority color, not the average.

The German and English studies reported color names. Thus, for each color–letter association in English and German, we transformed the color name into its hue, saturation, and luminance using a large-scale internet color name survey (Munroe, 2010) and subsequent transformation from RGB to HSV color space.

Letter Frequencies. For English, letter frequencies were estimated using three sources. First, they were tallied from the words of the Oxford English Dictionary (OED). In this dictionary, E is the most frequent letter (11.2%) and Q is the least frequent letter (0.196%). Though the dictionary provides broad coverage of the English language, each word is included only once and thus the prevalence of words in writing is not taken into account. For example, "the" is the most common English word and is used in nearly every sentence, yet the OED has only a few entries for its various usages. The prevalence of "the" in writing could cause a synesthete to associate a lower frequency color with the letters "t," "h," and "e." Therefore, in a second analysis, we used letter frequencies from the Brown Corpus, an anthology of the English language that includes excerpts from English literature, speeches, and music. The Brown Corpus represents a more accurate cross-section of the printed language that a synesthete experiences in daily life (Roland et al., 2007). Lastly, because a synesthete's grapheme–color associations are formed in childhood (Eagleman, 2007), we compiled a corpus of ten popular children's books (Supplemental Information), which is perhaps a better representation of childhood experience with printed text.

For Spanish and German, letter frequencies were extracted from cryptographic textbooks that compiled these frequencies (Pratt, 1942; Beutelspacher, 2005).

Correlation between letter frequency and hue. Letters associated with brown and gray were excluded from analysis because brown and gray have no clearly defined hue (For Spanish: B, H; For English: D, U, Z; For German: D, H, K, L, W). Per standard procedure, the letters "O" and "I" were omitted because synesthetes tend to confuse them with digits "0" and "1", which have their own associations (Barnett et al., 2008).

The reported correlations between color and letter frequency are circular–linear correlations between hue and log letter frequency. Hue is specified as an angle on the color wheel and is thus a circular dimension. Effects of frequency are most commonly measured in log units (see, e.g., Zipf's law) (Zipf, 1935).

Results

The frequency of a letter was correlated with the hue of the color most commonly associated with it across English, Spanish, and German (*t*-test with Fisher-transformed correlation coefficients, mean $r = 0.64$, $t(2) = 6.9$, $p = 0.020$; Figs. 1 and 2, Table 1). For English-speaking synesthetes, the correlations were uniformly strong with letter frequencies derived from the dictionary ($r = 0.69$), from children's books ($r = 0.67$), and from the Brown Corpus ($r = 0.69$). This is likely because the letter frequencies do not differ much across the different sources — the average correlation between the letter frequencies across the three sources is $r = 0.98$.

English	E T A N S H R L C M W F G Y P B V K J X Q Z
Spanish	E A S R N D L T C M U P G V Y Q F J Z X W K
German	E N S R A T U G C M B F Z V P J Y X Q

Figure 1. Letter frequencies and associated colors. Letters are ordered from most frequent (left) to least frequent (right) and are shaded in the hue most commonly associated with them. For example, nearly every German synesthete sees “N,” one the most common letters in the German language, as red (Emrich, Schneider, Zedler, 2004). On the other hand, most English-speaking synesthetes see “N," a moderately frequent letter in English, as green. (Some letters are excluded because they are associated with a color of indeterminate hue, such as brown or gray.)

Table 1. Circular correlations between hue and letter frequency.

	English hues	Spanish hues	German hues
English letter frequencies	***r* = 0.69** ***p* = 0.0069**	*r* = 0.40 *p* = 0.16	*r* = 0.64 *p* = 0.016
Spanish letter frequencies	*r* = 0.58 *p* = 0.030	***r* = 0.50** ***p* = 0.056**	*r* = 0.58 *p* = 0.036
German letter frequencies	*r* = 0.52 *p* = 0.060	*r* = 0.25 *p* = 0.49	***r* = 0.71** ***p* = 0.0068**

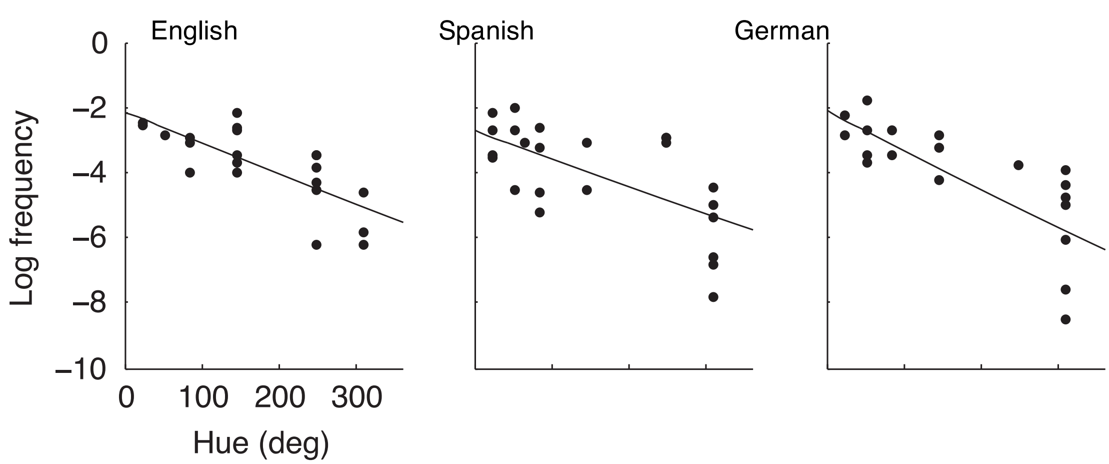

Figure 2. Circular correlations between hue and letter frequency. Plots depict circular–linear correlations between hue and log frequency.

The observed relationship between letter frequency and hue does not derive from high-frequency letters being associated with the most common color terms. To test this, we measured the partial correlation between hue and letter frequency controlling for color term frequency. Color-term frequencies for each color in each language were determined by their frequency in printed text (Google Ngram viewer). The correlation between letter frequency and color hue held across the three languages (*t*-test with Fisher-transformed correlation coefficients, mean $r = 0.63$, $t(2) = 7.7$, $p = 0.017$: Table 2).

Table 2. Partial correlations between color hue and letter frequency, controlling for color term frequency.

	English	Spanish	German	(Pooled)
r	0.70	0.50	0.67	0.63
p	0.0055	0.054	0.012	0.017

Because there is a small correlation between a letter's frequency and its position in the alphabet (English: $r = -0.29$; Spanish: $r = -0.36$; German: $r = -0.32$), we computed the partial correlation between color hue and letter frequency controlling for letter order (Table 3). Across the three languages, there is strong correlation even controlling for letter order (*t*-test with Fisher-transformed correlation coefficients, mean $r = 0.61$, $t(2) = 7.3$, $p = 0.018$).

Table 3. Partial correlations between color hue and letter frequency, controlling for letter order.

	English	Spanish	German	(Pooled)
r	0.67	0.47	0.66	0.61
p	0.0095	0.079	0.013	0.018

Lastly, we performed a partial correlation between color hue and letter frequency, controlling for both letter order and color term frequency (*t*-test with Fisher-transformed correlation coefficients, mean $r = 0.60$, $t(2) = 5.6$, $p = 0.030$: Table 5). The correlation between color hue and letter frequency is not due to color term frequency and letter order in the alphabet.

Table 5. Partial correlations between color hue and letter frequency, controlling for color term frequency and letter order.

	English	Spanish	German	(Pooled)
r	0.71	0.44	0.61	0.60
p	0.0047	0.11	0.025	0.030

Interim Discussion

We found a correlation between the frequency of a letter in language and the hue that synesthetes associate with it across three languages and three distinct groups of synesthetes. We considered other possible influences, such as letter order and color term frequency, and found that the correlation between frequency and hue held when eliminate the contribution of these other factors. This finding helps to explain the cause of individual grapheme–color associations and suggests that synesthetes' exposure to text guides the grapheme–color associations that they form.

Experiment 2

Next, we asked whether the frequency–hue correlation observed in synesthetes is specific to synesthetic associations, or if it is a general property of semantic associations between letters and colors among all observers. To address this question, we asked a group of non-synesthetic participants to report the color they most strongly associated with each letter of the English alphabet. If the frequency–hue correlation observed in Experiment 1 is a general property, then the same correlation should be observed for this non-synesthetic group.

Methods

Participants. Two hundred and fifty deidentified non-synesthetes were asked to assign a color to each letter of the English alphabet. Participants were recruited from Amazon's Mechanical Turk, an online marketplace where people perform short tasks for pay (Buhrmester, Kwang, & Gosling, 2011).

Stimulus and Procedure. On each trial, the participant saw a capital letter from the English alphabet, presented in a random order. The participant reported the color most closely associated with that letter in the form of a written color name. (This was done to match the procedure used in Experiment 1.) Each color name was converted to an RGB value, based on a large-scale internet survey (Monroe, 2010), which determined the RGB values of 954 common color names as defined by several hundred thousand participants. The RGB value was then converted to HSV.

Results

The average circular correlation between letter frequency and letter hue among the non-synesthetes was $r = 0.28 \pm 0.01$ (± SE), half the correlation found in the synesthetic population and significantly less (Steiger's test for dependent correlations with a common variable, $z = 2.58$, $p = 0.0099$). Many of the non-synesthetes used a name-matching strategy, in which they reported a color whose first letter matches the target letter (e.g. F is fuchsia, I is indigo). However, others reported color–letter associations that were not governed by the first letter of the paired color name. Perhaps these two strategies lead to different strengths of letter frequency–color hue correlations. Therefore, we examined these populations (“matching” vs. “non-matching”) separately. To discriminate between them, for each participant, we counted the number of letters (out of 26) associated with colors that begin with that letter (“matching”). We found no meaningful correlation between the amount of matching and the strength of the color-letter associations that are formed ($r= 0.04$, $p= 0.53$). The average correlation between hue and frequency was the same for participants who matched all 26 letters (0.29 ± 0.03) as it was for participants who did not (0.28 ± 0.01). Thus, the correlation between letter frequencies and color hues for non-synesthetes was half as strong as for synesthetes, no matter which strategy the non-synesthetes used.

Discussion

The correlation between letter frequency and associated color hue demonstrated in Experiment 1 is twice as strong in synesthetes as in non-synesthetes, which suggests that the processes that lead to reported grapheme–color associations differ between the groups.

General Discussion

We found that synesthetes' exposure to written language guides the grapheme–color associations that they form. Specifically, we found that among English-, Spanish-, and German-speaking synesthetes, the hue associated with a letter is strongly correlated with the frequency of that letter in written language. Letters that are more frequent are associated with red, orange, and yellow; less frequent letters are associated with green, blue, and violet. Conversely, when asked to assign a color to each letter, non-synesthetes assigned colors to letters in a different pattern, showing a weak correlation between letter frequency and hue and often assigning colors by semantic and orthographic association (e.g., A/apple/red or Y/yellow).

Because synesthesia has a genetic component (Asher et al., 2009), previous studies have assumed that the colors associated with each letter are genetically determined and specific to each synesthete, such that even fraternal twin pairs who shared a common childhood experience differ in their associations (Smilek, Dixon, Cudahy, & Merikle, 2001; Smilek, Dixon, & Merikle, 2005). Others suggested that color associations are governed by childhood experiences, such as playing with colored alphabet kitchen magnets (Witthoft & Winawer, 2005). The current results suggest that the colors associated with each letter are affected by the frequency with which the synesthete encounters letters in day to day life, rather than being innate or simply adopted from a single colored alphabet within the home. Presumably, the frequency of these letters within the synesthete's native alphabet is a proxy for the frequency of exposure to those letters.

Neural mechanisms

How would the frequency of exposure to letters influence the color associated with that letter? Here, we review the neural mechanisms involved in color and word processing to develop a plausible account of how color–grapheme pairings might arise through the interaction of color processing and word-form processing in the cortex.

Area V4 is located in the extrastriate cortex and plays a role in color perception, as evidenced by neuropsychological and neurophysiological data (Wade, Brewer, Rieger, & Wandell, 2002). Single cell recordings in monkeys have shown that neurons in V4 are clustered in columns by color selectivity (Kotake, 2009). Specifically, color discrimination index values, which compare the firing rates between a neuron's most and least preferred colors, are positively correlated between nearby neurons, but not between distant neurons, demonstrating that V4 neurons are clustered by ability to discriminate color (Kotake, 2009). Further, these clusters appear to be organized by long, middle, and short wavelength cone signals (Kotake, 2009), with the majority of V4 neurons preferring red or long-wavelength colors.

A second neural area relevant to synesthesia is the visual word form area (VWFA) in the temporal lobe. The VWFA specializes in the processes that underlie reading: recognizing, processing, and interpreting letters and words (Cohen & Dehaene, 2004).

Importantly, the VWFA is adjacent to V4. The two areas are simultaneously activated in synesthetes and there is evidence that they are linked (Brang, Hubbard, Coulson, Huang, & Ramachandran, 2010; Hubbard, Arman, Ramachandran, & Boynton, 2005).

Taking these features into account, neurophysiological models propose that synesthesia is driven by connectivity between the visual word form area and V4 (Brang et al., 2010; Hubbard et al., 2005). Within this framework, the VWFA receives biased inputs, with some letters more frequent than others (depending on the language), and area V4 has biased color selectivity, with more neurons responding to long-wavelength colors like red (which is presumably independent of language).

We propose that these two biases jointly determine which letters and colors become associated in synesthetes. For example, if associations were initially formed randomly between letters and colors, the more frequent letters would form more associations with long-wavelength neurons, and would therefore become associated with red. But this random association model would also predict that frequent letters would become associated with other colors, and that less frequent letters would also become strongly associated with red. To explain why that does not occur, we must assume that, as some associations are reinforced through experience, others are inhibited. For instance, as frequent letters become strongly associated with red, their association with other colors must be inhibited, and the association between less frequent letters and red must be inhibited as well. According to this frequency-based-inhibition model, greater letter frequency and greater frequency of color selectivity combine to strongly associate higher frequency letters in word-processing regions (e.g., VWFA) with more prevalent colors in color-processing areas (e.g., V4).

While this account is speculative, it is consistent with known physiology, and provides a plausible mechanism for the observed relationship between letter-frequency and letter-hue associations across different languages.

Conclusion

Among English-, Spanish-, and German-speaking synesthetes, the hue associated with a particular letter is correlated with that letter's frequency in text. The correlation between color hue and letter frequency in a non-synesthete control population was weaker, supporting previous claims that synesthesia is a distinct neurological phenomenon. We propose that synesthetic associations are formed between word processing areas (e.g., VWFA) and color processing areas (e.g., V4), with binding between letters and colors depending jointly on biases in the input (i.e., letter frequency, which varies with language) and biases in color processing (i.e., the uneven distribution of color selectivity within V4, which is independent of language). Though synesthesia has a genetic component (Asher et al., 2009), the precise letter–color pairings that are formed are affected by perceptual experience.

Acknowledgements

Many thanks to Jennifer Gordinier for her help in developing this project. Thanks to Paola Torres for help with translation. The German synesthete data was used with permission from Dr. Markus Zedler of the Hannover Medical School in Hannover, Germany. It is published in his book *Welche Farbe hat der Montag? (What Color is Monday?).*

References

Asher, J. E., Lamb, J. A., Brocklebank, D., Cazier, J. B., Maestrini, E., Addis, L., ... & Monaco, A. P. (2009). A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. *The American Journal of Human Genetics*, *84*(2), 279–285.

Bargary, G. (2007). *Illusory multisensory interactions in synaesthesia*. University of Cambridge.

Barnett, K. J., Finucane, C., Asher, J. E., Bargary, G., Corvin, A. P., Newell, F. N., & Mitchell, K. J. (2008). Familial patterns and the origins of individual differences in synaesthesia. *Cognition, 106*, 871–893.

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

Beeli, G., Esslen, M., & Jancke, L. (2007). Frequency correlates in grapheme–color synesthesia. *Psychological Science*, *18*(9), 788–792.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Beutelspacher, Albrecht (2005). *Kryptologie* (7 ed.). Wiesbaden: Vieweg. p. 10.

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.

Brown University Corpus Letter Frequencies per 10,000 Letters. *Index of Coincidence and Polyalphabets*.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Cohen, L. & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, *22*(1), 466–476.

Courchesne, E., Pierce, K., Schumann, C. M., Redcay, E., Buckwalter, J. A., Kennedy, D. P., & Morgan, J. (2007). Mapping early brain development in autism. *Neuron*, *56*(2), 399–413.

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.

Emrich, H., Schneider, U., & Zedler, M. (2004). *Welche Farbe hat der Montag?* Berlin: S. Herzel Verlag.

Feinberg, I. (1983). Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *Journal of Psychiatric Research*, *17*(4), 319–334.

Hancock, P. (2005). Monozygotic twins' colour-number association: A case study. *Cortex, 42*(2), 147–150.

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.

Kotake, Y. (2009). Organization of color-selective neurons in macaque visual area V4. *Journal of Neurophysiology, 102*, 15–27.

Nunn, J. A., Gregory, L. J., Brammer, M., Williams, S. C. R., Parslow, D. M., Morgan, M. J., Morris, R. G., Bullmore, E. T., Baron-Cohen, S., & Gray, J. A. (2002) Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. *Nature Neuroscience, 4*, 371–375.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*(3), 303.

Munroe, R. (2010). Xkcd color name survey. Retrieved from https://xkcd.com/color/rgb/.

Pratt, Fletcher (1942). *Secret and Urgent: the Story of Codes and Ciphers*. Garden City, N.Y.: Blue Ribbon Books, 254–5.

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, 53–84.

Rissman, J. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, 8, 1160–1175.

Scullion, V. & Treby, M. (2010). Creative synaesthesia in E. T. A. Hoffmann's Ritter Gluck. *European Review*, 18, 239–262.

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*, 1069–1085.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., ... & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*(8), 1024–1033.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2001). Synaesthesia: A case study of discordant monozygotic twin. Neurocase, 8, 342–383.

Smilek, D., Dixon, M. J., Cudahy, C., & Merikle, P. M. (2002). Synesthetic color experiences influence memory. *Psychological Science*, *13*(6), 548–552.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2005). Synaesthesia: Discordant male monozygotic male twins. *Neurocase*, 11, 363–370

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1470), 979–983.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*, 348–379.

Wade, A. R., Brewer, A. A., Rieger, J. W., & Wandell, B. A. (2002). Functional measurements of human ventral occipital cortex: retinotopy and colour.*Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1424), 963–973.

Ward, J. & Simner, J. (2003). Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition, 89*(3), 237–261.

Wei, T. (2012). Predicting conceptual processing capacity from spontaneous neuronal activity of the visual word form area. *Journal of Neuroscience*, 32, 481–489.

Witthoft, N. & Winawer, J. (2005). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex*, *42*(2), 175–183.

Zipf, G. K. (1935). The psycho-biology of language.

Supplemental information

The following children's books were used to derive letter frequencies:

Where the Wild Things Are

Goodnight Moon

Green Eggs and Ham

The Very Hungry Caterpillar

The Giving Tree

Oh! The Places You'll Go

Where the Sidewalk Ends

The Cat in the Hat

The Velveteen Rabbit

Harold and the Purple Crayon

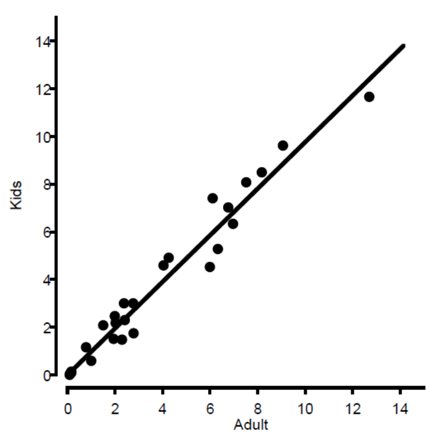

Figure S1. Plot of Brown corpus ("Adult") letter frequencies compared to children's books ("Kids") letter frequencies.