CONSISTENCY AND STRENGTH OF GRAPHEME-COLOR ASSOCIATIONS ARE SEPARABLE ASPECTS OF SYNESTHETIC EXPERIENCE

Simon Lacey[1,2,4], **Margaret Martinez**[4,*,†], **Nicole Steiner**[**,5,†], **Lynne C. Nygaard**[5], **K. Sathian**[1,2,3,4,5]

[1]Department of Neurology, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA.

[2]Department of Neural & Behavioral Sciences, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA.

[3]Department of Psychology, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA.

[4]Department of Neurology, Emory University, Atlanta, GA 30322, USA

[5]Department of Psychology, Emory University, Atlanta, GA 30322, USA

Abstract

Consistency of synesthetic associations over time is a widely used test of synesthesia. Since many studies suggest that consistency is not a completely reliable feature, we compared the consistency and strength of synesthetes' grapheme-color associations. Consistency was measured by scores on the Synesthesia Battery and by the Euclidean distance in color space for the specific graphemes tested for each participant. Strength was measured by congruency magnitudes on the Implicit Association Test. The strength of associations was substantially greater for synesthetes than non-synesthetes, suggesting that this is a novel, objective marker of synesthesia. Although, intuitively, strong associations should also be consistent, consistency and strength were uncorrelated, indicating that they are likely independent, at least for grapheme-color synesthesia. These findings have implications for our understanding of synesthesia and for estimates of its prevalence since synesthetes who experience strong, but inconsistent, associations may not be identified by tests that focus solely on consistency.

Corresponding author: K. Sathian, Department of Neurology, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA, Tel: 717-531-1801, Fax: 717-531-0384, ksathian@pennstatehealth.psu.edu.
*Present address: Rowan School of Osteopathic Medicine, Stratford, NJ.
**Present address: University of Texas at Austin, Austin, TX
†These authors contributed equally to the present work.

1. INTRODUCTION

Synesthesia is a phenomenon in which ordinary stimuli, such as visual letters or auditory tones ('inducers'), elicit involuntary and unrelated secondary experiences, such as colors or shapes ('concurrents': Novich et al., 2011; Simner, 2012; Ward, 2013). These synesthetic associations have several diagnostic characteristics: for example they are normally considered to be unidirectional (but see Weiss et al., 2008; Gebuis et al. 2009, 2009a; Shalgi & Foxe, 2009) in addition to being involuntary, and they are consistent over time. Consistency has been described as the 'hallmark' of synesthesia (Akiva-Kabiri et al., 2014; Ovalle Fresa & Rothen, 2019) and tests of consistency are considered the 'gold standard' (Carmichael et al., 2015; Ovalle Fresa & Rothen, 2019) in determining whether an individual is synesthetic or not. Tests of consistency are employed in instruments that examine multiple kinds of synesthesia, for example, the Test of Genuineness (Revised: TOG-R, Asher et al., 2006), and the online Synesthesia Battery (SB: Eagleman et al., 2007; Carmichael et al., 2015) as well as instruments that only test a single kind, for example, pitch-space (Linkovski et al., 2012), timbre-color (Menouti et al., 2015), sequence-space (Rothen et al., 2016), and lexical-gustatory (Ipser et al., 2020) synesthesia. Such instruments have tested consistency in both the short term, over approximately 30 minutes to several hours (Eagleman et al., 2007; Menouti et al., 2015; Rothen et al., 2016; Ipser et al., 2020), and the long term, over months to years (Asher et al., 2006; Linkovski et al., 2012) or even decades (Simner & Logie, 2007).

However, despite its ubiquity in tests for synesthesia, whether consistency is truly a central feature has been questioned (Simner, 2012; Eagleman, 2012). Synesthetic associations are known to be variable during childhood although they are mostly settled by age 10 (Simner & Bain, 2013). But they may also vary during adulthood (Niccolai et al., 2012, Meier et al., 2014; Simner et al., 2017; Chromý et al., 2019), and even according to mood (Kay et al., 2015). Both short-term (i.e., within-test) and long-term (i.e., test-retest) consistency may vary radically over the course of approximately 18 months during adulthood (Chromý et al., 2019). The changes in associations are many and various: the range of synesthetic experiences of an individual can expand to involve a different category of inducers for the same concurrent or vice versa and the self-rated intensity of the concurrent experience can increase or decrease (Niccolai et al., 2012). In grapheme-color synesthesia, the number of graphemes with color associations may increase or decrease (Niccolai et al., 2012; Chromý et al., 2019) and there can be radical changes in the color associated with a grapheme, e.g., from 'light salmon' to 'dark sea green' or from gray to red (Niccolai et al., 2012; Chromý et al., 2019). In addition, bright colors (but not necessarily brightness itself) may be less frequently experienced with age with darker, more subdued colors occurring more commonly (Meier et al., 2014). Luminance and saturation decrease with increasing negative mood and anxiety (Kay et al., 2015) and with age (Simner et al., 2017). (Note that a recent study of both Latin alphanumeric graphemes and Japanese characters suggests that

consistency may be stable over both the short term (1–10 months) and long term (5–8 years) but varied with the rated familiarity of the grapheme in that less familiar graphemes were also less consistent over both the short- and long-term [Uno et al., 2021]).

The unreliability of consistency as a marker for synesthesia has important implications for our phenomenological understanding of synesthesia and its identification. The fact that synesthetic color associations desaturate over time means that many older grapheme-color synesthetes may fail the consistency test (Simner et al., 2017) and there may be individuals who, while definitely having specific color associations for particular graphemes, nonetheless experience different colors at different times and thus exhibit low consistency (Simner, 2012). Thus, it is valuable to examine other potential characteristics of synesthesia, such as the strength of associations, in addition to their consistency. Association strength has been studied in relation to the crossmodal correspondence between auditory pitch and visual shape using a statistical learning paradigm (Chan & Dyson, 2015), but that study did not test synesthetes. In a case study of a grapheme-color synesthete, Hubbard et al. (2006) assessed association strength by asking the synesthete to rate it on a Likert scale, but this method is subjective. Here, we investigated both the strength and consistency of associations in grapheme-color synesthesia. We used the Implicit Association Test (IAT: Greenwald et al., 1998) to provide an objective measure of strength. Consistency was evaluated using scores on the Synesthesia Battery (SB: Eagleman et al., 2007), which reflect the variability of the colors selected to match graphemes on three trials separated in time, and an analogous measure, the Euclidean distance between coordinates in RGB color space for the colors chosen on those repeated trials. The IAT has previously been used to examine crossmodal correspondences (Parise & Spence, 2012; Anikin & Johansson, 2019) and the universal color-shape associations proposed by the artist Kandinsky (Makin & Wuerger, 2013). We have previously used the IAT to compare synesthetes to non-synesthetes in their processing of crossmodal correspondences (Lacey et al., 2016) but, to our knowledge, the present study is the first to use the IAT to examine the strength of synesthetic associations. Intuitively, one might reason that consistency should arise from strong synesthetic associations and therefore that strength and consistency would be significantly positively correlated. Alternatively, consistency and strength of association may reflect different aspects of synesthetic correspondences and therefore may be weakly or not correlated.

2. MATERIALS & METHODS

2.1 Participants

Thirty-six people (6 males, 30 females; mean age 21 years) took part in this study: 18 synesthetes (identified based on their SB scores, see below) and 18 age- and gender-matched non-synesthetic controls (3 males, 15 females in each group; mean age: synesthetes 21 years 4 months, non-synesthetes, 20 years 11 months; the ages did not differ significantly [t_{34} = .4, p = .7]). All participants gave informed consent and were compensated for their time. All procedures were approved by the Emory University Institutional Review Board.

During data collection, the SB was offline while being transferred to a new server and some participants performed a substitute task (detailed below) and only completed the SB

later. One synesthete who did the substitute task did not complete the SB. We also had technical difficulty accessing SB data for one other synesthete; for this person, we were able to ascertain only their different synesthesias and their grapheme-color SB score.

For participants with sufficient SB data, SB scores confirmed that all 17 synesthetic participants experienced grapheme-color synesthesia. Grapheme-color synesthesia was the only (1) or primary (8) synesthesia for 9 of these synesthetes by reference to their SB scores (it remains possible that their actual primary synesthesia was one not tested by the SB). Sixteen synesthetes reported that they experienced at least one other synesthesia, including one who reported color associations for Greek alphabet graphemes. The remaining additional synesthesias predominantly involved color concurrents, the inducers being weekdays, months (11 each); personalities (5); the sounds of different musical instruments, emotions, odors (4 each); musical chords, touch, taste, temperature (3 each); pain (2); and musical key signature, pitch, and orgasm (1 each). Other types of synesthesia were also reported by our participants: 7 people had sequence-space synesthesia, 2 reported grapheme-personality associations, and one indicated that languages and concepts involved colors, shapes, and textures. Synesthetes are divided into 'projectors' who see visual concurrents "out in the world" and 'associators' who "see" visual concurrents in the mind's eye or who have a strong feeling for an association (Dixon et al., 2004). Due to technical problems with the SB after it was moved to a new server, we were only able to obtain projector/associator scores for 8 of our synesthetes; these revealed only one projector, consistent with findings that the associator sub-type is more common (Dixon & Smilek, 2005).

2.2   Synesthesia Battery

Participants who claimed to experience grapheme-color synesthesia completed the SB (Eagleman et al., 2007) in order to verify their synesthetic status. The SB reliably identifies certain common varieties of synesthesia (Carmichael et al., 2015). For grapheme-color synesthesia associated with the Latin alphabet and Arabic numerals, all 36 graphemes are presented three times in random order and the participant uses a color-picker to select the best match for the color they experience for that grapheme. Color-picker responses are converted to a single SB score. An individual is considered a synesthete if their SB score is less than 1 and a non-synesthete if their score is more than 2; where an individual's score falls between 1 and 2, their synesthetic status cannot be reliably determined (Eagleman et al., 2007).[1] The SB also records the RGB values for the colors chosen on each of the three trials for each grapheme. RGB values range from 0 to 255 for each component in arbitrary units of intensity, and colors are defined in RGB triplets: black and white are respectively produced by zero (0,0,0) and maximal (255,255,255) intensity in all components (Hunt, 2003). Consistency was measured by calculating the mean Euclidean distance between the three color estimates for each grapheme (i.e., the mean of the 1st vs 2nd, 1st vs 3rd, and 2nd vs 3rd). The Euclidean distance is given by:

---

[1]All seventeen synesthetes who completed the SB scored less than 1 for grapheme-color and their additional synesthesias reported above, but seven of them reported additional synesthesias with SB scores that fell in the indeterminate range (two reported pitch-color but scored 1.05 and 1.63; two reported weekday-color, scoring 1.0 and 1.13; and three reported month-color, scoring between 1.11 and 1.24). Furthermore, of eight people who volunteered because they thought they experienced grapheme-color synesthesia but who were excluded on the basis of their SB score, seven had grapheme-color SB scores in the indeterminate range (scores ranged from 1.09 to 1.67) and only one was clearly a non-synesthete (scoring 2.4).

$$\sqrt{\left[(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2\right]} \qquad \text{Equation 1:}$$

While the SB was offline, we devised a substitute task using the online color-picker at https://www.rapidtables.com/web/color/RGB_Color.html. We chose an RGB color-picker in order to match the SB, which also uses this color space (but see also Rothen et al., 2013 for the use of online color-pickers). This enabled participants who self-reported as grapheme-color synesthetes to choose their synesthetic colors in the same way as the SB, i.e. they could select the approximate color using a sliding bar and then adjust this more precisely by clicking within the resulting color square and checking their selection in the resulting color patch. Visually, the SB and online color-picker were very similar, employing the same spatial layout. The only differences were: a black background in the SB, as opposed to white in the online task; the SB had an additional slider bar to make fine adjustments for hue; the SB recorded RGB values automatically whereas these had to be transcribed from the display in the online task. As with the SB, participants were asked to select their synesthetic color for each grapheme three times in random order. Mirroring the SB, which changed the starting color on each letter presentation, the experimenter manually re-set the picker to a different color on each trial to avoid the use of memory strategies and spatial cues. RGB values were recorded for each trial, and the Euclidean distance between the three estimates was determined as above.

14 synesthetes completed the SB and 4 completed the online color-picker task at the time of testing. Whether participants completed the SB or the online color-picker, we extracted the RGB values and chose the two graphemes with the most consistent color choices (i.e., the shortest Euclidean distances) that were also easily distinguishable from each other, as confirmed with each synesthete on an individual basis, because individual differences in the synesthetic color palette forming each synesthete's associations meant that the range of colors to choose from was more limited for some synesthetes (see 2.4 for the mean Euclidean distance in RGB space for the two colors tested). Because we presented all graphemes in the IAT in black on a white background, we avoided graphemes with associations to black and white. Three of the four participants who chose their synesthetic colors using the online color-picker completed the SB when it became available again in order to confirm their synesthetic status using a standardized test.

Age- and gender-matched control participants were recruited from volunteers who understood the concept of synesthesia and knew that they did not experience it. These participants did not take the SB, but their non-synesthetic status was reviewed and confirmed in a screening interview that covered the definition of synesthesia, a range of different inducers and concurrents, and the projector/associator sub-types.

2.3 Implicit Association Test

The IAT was presented using a procedure we have previously described (Lacey et al., 2016). Presentation software (Neurobehavioral Systems Inc., Albany CA) was used to run the experiment and record response times (RTs). Participants were instructed to associate pairs of stimuli with one of two response keys (the 'left' and 'right' arrow keys on a normal US

'QWERTY' keyboard). The pairs always consisted of one grapheme and one color and, in separate blocks of trials, were either congruent or incongruent. Figure 1B shows a fictitious example in which synesthetically congruent pairs are "A"/red (both to be associated with the left arrow key) and "B"/blue (associated with the right arrow key) and the incongruent pairs are "A"/blue (left arrow key) and "B"/red (right arrow key). However, each stimulus was presented in isolation, i.e. a trial consisted either of a grapheme ("A" or "B") or a color (red or blue) and participants were asked to respond by pressing the assigned key as quickly as possible.

Following the procedure described in Lacey et al. (2016), each synesthete's pair of grapheme-color associations was tested in two runs. In each run there were 96 trials divided into a block of 48 congruent trials followed by a block of 48 incongruent trials, or vice versa, giving a total of 192 active trials across the two runs. Before each block of 48 trials, participants read on-screen instructions describing the response key associations to be used and then performed 12 practice trials with on-screen feedback as to accuracy (practice trials were not included in the analysis and feedback was only given for the practice trials). Within each run, half the trials were graphemes and half were colors, split evenly between congruent and incongruent trials. Trials consisted of a blank 1000ms followed by either a grapheme or a color stimulus for 1000ms and were terminated either by the participant pressing a response key or automatically 3500ms after stimulus onset. The length of each active block thus varied between participants but was a maximum of 330s. RTs were measured from stimulus onset. Age- and gender-matched control participants took the IAT with the same grapheme-color pairs as the synesthete to whom they were matched. The order of the two runs was counterbalanced across participants.

2.4 Procedure

Both the SB and the online color-picker had three estimates of the synesthetic color for each grapheme and there was no way of knowing which of these was closest to the synesthete's actual experience. Therefore, on the day of testing, participants were instructed to choose the most accurate color by making a single choice with the online color-picker for each of the two graphemes for the IAT. The mean Euclidean distance between the color associations for the two graphemes tested was 222.9 (standard deviation: ±69.2) intensity units. As noted in 2.2, values for each component in RGB space range from 0 to 255, thus a mean Euclidean distance of 222.9 indicates that the two colors were far apart in RGB space and therefore easily distinguishable.

Accuracy (% correct) was based on those trials for which a response was recorded and the magnitude of the congruency effect was given by:

$$[(\text{accuracy on congruent trials} - \text{accuracy on incongruent trials})/(\text{accuracy on congruent trials} + \text{accuracy on incongruent trials})] \times 100 \qquad \text{Equation 2:}$$

RT analyses were based on correct trials only and excluding any trials with RTs more than 3 standard deviations away from the individual mean, calculated separately for each run. The magnitude of the congruency effect for RTs was given by:

$$[(\text{incongruent RT} - \text{congruent RT})/(\text{incongruent RT}+\text{congruent RT})] \times 100 \qquad \text{Equation 3:}$$

Data were analyzed using IBM SPSS v27 (IBM Corporation, Armonk NY) and effect sizes (Cohen's d) were calculated using the online tool provided by Lenhard & Lenhard (2016). As noted in the Introduction, consistency might be independent of strength and therefore these two attributes of synesthesia might not be significantly correlated. In fact, we did not find a correlation between these two aspects of synesthesia. This involves accepting the null hypothesis and we therefore calculated Bayes factors (BF) in order to assess the strength of the evidence for the null hypothesis (Dienes, 2011; Cumming, 2014). SPSS computes the ratio of the null hypothesis to the experimental hypothesis and, for Pearson correlations, we accepted the default setting of the JZS (Jeffreys-Zellner-Siow) Bayes factor (Jeffreys, 1961; Zellner & Siow, 1980). We interpret BFs of 1–3 as indicating weak, and 3–10 as indicating substantial, evidence for the null hypothesis. This largely follows Jeffreys (1961); however, we acknowledge that there are several interpretive scales differing in both cut-off points and the rhetorical force of the descriptive term (Jarosz & Wiley, 2014; Hoijtink et al., 2016). Finally, note the degrees of freedom for correlations are given by N-2.

3. RESULTS

3.1 Group differences

Trials for which there was no response accounted for 0.5% of total trials. Correct responses were recorded for 94.5% of the remaining trials; for RT analyses, 1.9% of correct response trials were excluded as outliers.

3.1.2 RTs—Repeated-measures ANOVA (RM-ANOVA) of RTs (between-group factor: synesthetic status – synesthetes, non-synesthetes; within-group factor: trial type – congruent, incongruent) showed a main effect of trial type ($F_{1,34} = 42.6$, p < .001, d = 2.2) in which RTs for congruent trials were faster (mean ± sem, 505 ± 20ms) than for incongruent trials (678 ± 37ms). There was no main effect for synesthetic status ($F_{1,34} = 1.2$, p = .3, d = .4) but, crucially, there was an interaction between synesthetic status and trial type ($F_{1,34} = 20.9$, p < .001, d = 1.6: Figure 2A). Post-hoc t-tests (Bonferroni-corrected α for four tests = .0125) showed that RTs for synesthetes were significantly faster for congruent (471 ± 11ms) compared to incongruent (766 ± 44ms) trials ($t_{17} = -7.2$, p < .001, d = 1.9) whereas for non-synesthetes there was no significant difference (congruent, 539 ± 37ms; incongruent, 591 ± 55ms: $t_{17} = -1.5$, p = .1, d = .2). While there was no significant difference between synesthetes and controls for either congruent ($t_{34} = -1.7$, p = .09, d = .6) or incongruent ($t_{34} = 2.5$, p = .02, d = .8) trials, synesthetes were nearly 30% slower than controls on the incongruent trials. This interaction effect suggests that synesthetes have difficulty inhibiting their associations on incongruent trials but only a small (non-significant in the present data set) advantage for congruent trials (although there is a limit on how fast a response can be and thus facilitatory effects might well be compressed in the congruent trials). This pattern of results did not change if we excluded the synesthete who did not complete the SB and their matched control.

3.1.3   RT congruency magnitudes—We also performed RM-ANOVA of congruency magnitudes for RTs (between-group factor: synesthetic status – synesthetes, non-synesthetes; within-group factor: trial type – grapheme, color). There was a large main effect of synesthetic status ($F_{1,34} = 31.3$, $p < .001$, $d = 1.9$) in which RT congruency magnitudes were greater for synesthetes (22.5 ± 2.5) than non-synesthetes (3.8 ± 2.2). There was a main effect of trial type ($F_{1,34} = 9.1$, $p = .005$, $d = 1.0$) in which RT congruency magnitudes were greater for color trials than grapheme trials (14.3 ± 2.4 vs 11.9 ± 2.2 respectively). There was a significant interaction between synesthetic status and trial type ($F_{1,34} = 4.7$, $p = .04$, $d = .7$: Figure 2B) in which RT congruency magnitudes were greater for color (24.4 ± 2.7) than grapheme (20.4 ± 2.4) trials for synesthetes (Bonferroni-corrected α for four tests = .0125: $t_{17} = -3.1$, $p = .007$, $d = .4$) but not for controls (color, 4.1 ± 2.2; grapheme, 3.5 ± 2.3: $t_{17} = -.8$, $p = .5$, $d = .07$). Reflecting the main effect, RT congruency magnitudes were significantly greater for synesthetes than controls for each trial type (color, $t_{17} = 5.1$, $p < .001$, $d = 1.7$; grapheme, $t_{17} = 5.8$, $p < .001$, $d = 1.9$). This pattern of results did not change if we excluded the synesthete who did not complete the SB and their matched control, except that the interaction narrowly missed significance ($F_{1,32} = 3.9$, $p = .055$, $d = .7$) although post-hoc tests showed the same pattern of results.

3.1.4   RT congruency magnitudes for color trials only—In a post-hoc analysis, we examined why synesthetes had larger RT congruency magnitudes for color trials than grapheme trials. If synesthesia is unidirectional, i.e., graphemes induce colors but not vice versa, congruency magnitudes might be expected to differ between grapheme and color trials. Given the directionality of grapheme-color synesthesia, the concurrent color induced by a grapheme might prime responses to a subsequent color trial if the colors were the same, or inhibit responses if they differed (see Mattingley et al., 2001). Additionally, in the IAT, all responses are made under conditions where response key associations are either congruent or incongruent; thus, an alternative explanation is that magnitudes are more affected by the congruency of response key associations than that of the preceding trial. To investigate whether the increased color congruency magnitudes occurred due to the preceding grapheme trial) or the response key associations, we analyzed all color trials that were immediately preceded by a grapheme trial (GC trials). We separated these into trials where the synesthetic color evoked by the grapheme was either congruent (GC_c) or incongruent (GC_i) with the real color presented in the color trial. We further subdivided these trials according to whether the response key associations for the color trial were congruent (R_c) or incongruent (R_i). The congruency magnitude for GC trials resulting from any influence of the preceding grapheme trial is given by:

$$[(R_cGC_i + R_iGC_i) - (R_cGC_c + R_iGC_c)/(R_cGC_i + R_iGC_i) + (R_cGC_c + R_iGC_c)] \times 100$$

Equation 4:

The congruency magnitude for GC trials resulting from any influence of the response key associations is given by:

$$[(\mathbf{R_i}GC_i + \mathbf{R_i}GC_c) - (\mathbf{R_c}GC_i + \mathbf{R_c}GC_c)/(\mathbf{R_i}GC_i + \mathbf{R_i}GC_c) + \mathbf{R_c}GC_i + \mathbf{R_c}GC_c)] \times 100$$

Equation 5:

RM-ANOVA of the GC trial congruency magnitudes showed a significant effect of synesthetic status ($F_{1,34} = 20.7$, $p < .001$, $d = 1.6$) in which synesthetes exhibited larger congruency magnitudes than non-synesthetes. There was a significant main effect in which congruency magnitudes were larger for response key associations (14.2 ± 2.5) than for the preceding grapheme trial ($1.3 \pm .8$: $F_{1,34} = 44.1$, $p < .001$, $d = 2.2$). Crucially, there was an interaction between synesthetic status and congruency magnitude type ($F_{1,34} = 30.3$, $p < .001$, $d = 1.9$: Figure 3). Post-hoc tests (Bonferroni-corrected α for four tests = .0125) showed that, for synesthetes, congruency magnitudes related to response key associations (24.5 ± 2.9) were larger than those arising from the preceding grapheme trial ($.9 \pm 1.1$: $t_{17} = 8.0$, $p < .001$, $d = 2.4$), whereas there was no difference for non-synesthetes (response key, 3.9 ± 2.3; preceding grapheme trial, 1.7 ± 1.3: $t_{17} = .9$, $p = .4$, $d = .3$). In addition, response key congruency magnitudes were larger for synesthetes than non-synesthetes ($t_{34} = 5.5$, $p < .001$, $d = 1.8$), whereas preceding grapheme trial magnitudes did not differ ($t_{34} = -.5$, $p = .6$, $d = .2$). This pattern of results suggests that, in fact, synesthetes' greater congruency magnitudes for color trials compared to grapheme trials (shown in Figure 2B) arose from within-trial effects with little interference/facilitation from the incongruence/congruence of the preceding grapheme trial.

3.1.5 Accuracy—RM ANOVA of accuracy (between-group factor: synesthetic status – synesthetes, non-synesthetes; within-group factor: trial type – congruent, incongruent) showed a main effect of trial type ($F_{1,34} = 16.4$, $p < .001$, $d = 1.4$) in which accuracy was greater for congruent (mean $\pm$ sem: $96.9 \pm .6\%$) compared to incongruent ($91.6 \pm 1.4\%$) trials. There was no main effect of synesthetic status ($F_{1,34} = .06$, $p = .8$, $d = .08$) and the interaction between synesthetic status and trial type only narrowly missed significance ($F_{1,34} = 3.9$, $p = .054$, $d = .7$: Figure 4A). Post-hoc t-tests (Bonferroni-corrected α for four tests = .0125) showed that synesthetes were significantly more accurate for congruent ($98.0 \pm .5\%$) compared to incongruent ($90.1 \pm 2.3\%$) trials ($t_{17} = 3.5$, $p = .003$, $d = 1.0$) whereas for non-synesthetes this comparison was not significant (congruent, $95.8 \pm 1.1\%$; incongruent, $93.1 \pm 1.4\%$: $t_{17} = 2.0$, $p = .06$, $d = .5$); there were no significant differences between synesthetes and controls for either congruent ($t_{34} = 1.7$, $p = .09$, $d = .6$) or incongruent ($t_{34} = -1.1$, $p = .3$, $d = .4$) trials. This pattern of results did not change if we excluded the synesthete who did not complete the SB and their matched control. Accuracy levels were overall quite high, so the absence of significant differences for some of the comparisons might reflect ceiling effects.

3.1.6 Accuracy congruency magnitudes—RM-ANOVA of congruency magnitudes for accuracy (between-group factor: synesthetic status – synesthetes, non-synesthetes; within-group factor: trial type – grapheme, color) showed that there was no main effect of synesthetic status ($F_{1,34} = 3.4$, $p = .07$, $d = .6$) although congruency magnitudes were larger overall for synesthetes (4.5 ± 1.5) than non-synesthetes ($1.4 \pm .8$). The main effect of trial type was not significant ($F_{1,34} = 1.0$, $p = .3$, $d = .3$) but the interaction between

synesthetic status and trial type only narrowly missed significance ($F_{1,34}$ = 3.9, p = .054, d = .7: Figure 4B). Post-hoc t-tests (Bonferroni-corrected α for four tests = .0125) showed that, between groups, congruency magnitudes for color trials were larger for synesthetes (6.3 ± 1.9) than non-synesthetes (1.0 ± .5: t_{34} = 2.7, p = .01, d = .9), whereas for grapheme trials there was no significant difference (synesthetes, 2.9 ± 1.4; non-synesthetes, 2.1 ± 1.6: t_{34} = .4, p = .7, d = .1); there were no within-group differences between grapheme and color magnitudes (synesthetes, t_{17} = −2.1, p = .049, d = .5; non-synesthetes, t_{17} = .7, p = .5, d = .2). This pattern of results did not change if we excluded the synesthete who did not complete the SB and their matched control.

3.2 Correlational analyses

3.2.1 RTs—There was no significant correlation (Pearson's r throughout) between consistency (grapheme-color SB scores) and strength (IAT congruency magnitudes for RTs: Figure 5A) across all trials (r_{15} = .06, p = .8, BF = 5.3) or when trial types were considered separately, for either grapheme (r_{15} = .1, p = .7, BF = 4.9) or color (r_{15} = .01, p = .9, BF = 5.4) trials (Bonferroni-corrected α for three tests = .017). Critically, the BFs for these analyses were all well over 3, indicating substantial support for the null hypothesis. (Note that there was also no significant correlation between consistency and strength in the analysis of GC trial congruency magnitudes relating to either response key associations, r = .01, p = .9, BF = 5.4, or the preceding grapheme trial, r = .4, p = .1, BF = 1.7). Due to the SB being offline, some synesthetes completed it later which might have potentially led to a different SB score than if this had been ascertained closer to the time of the IAT, but the results were similar if these individuals were excluded (all trials, r_{12} = −.16, p = .6, BF = 4.3; grapheme trials, r_{12} = −.1, p = .7, BF = 4.7; color trials, r_{12} = −.2, p = .5, BF = 4.0).

Moreover, the SB is a measure of overall consistency for all those graphemes for which an individual has associations while the IAT only tested associations for two graphemes. Therefore, we also measured consistency for the two graphemes used for the IAT by calculating the mean Euclidean distance in RGB space for the three color estimates (derived from RGB values extracted from either SB data or the online color-picker, whichever was completed immediately prior to performing the IAT) for the two graphemes tested and averaging across the pair. This allowed us to compare all synesthetes on the same footing, whether they completed the SB prior to the IAT or at a later date. However, there was still no significant correlation with congruency magnitudes for RTs, across all trials (r_{15} = .3, p = .2, BF = 2.2) or for grapheme (r_{15} = .2, p = .3, BF = 3.4) and color (r_{15} = .4, p = .1, BF = 1.5: Figure 6A) trials separately (Bonferroni-corrected α for three tests = .017); this was also the case for correlations with the mean Euclidean distances calculated for each grapheme separately.

3.2.2 Accuracy—Nor did we find a significant relationship between consistency and strength in terms of IAT congruency magnitudes for accuracy (Figure 5B) across all trials (r_{15} = .4, p = .07, BF = 1.1), or for either grapheme (r_{15} = .3, p = .2, BF = 2.2) or color (r_{15} = .4, p = .08, BF = 1.2) trials (Bonferroni-corrected α for three tests = .017). However, the BFs were in the 1–3 range, indicating weak support for the null hypothesis. As can be seen from Figure 5B, there was a potential outlier in terms of accuracy congruency magnitude,

but excluding this outlier did not change the significance of the results; in fact the BFs were now greater than 3, indicating substantial support for the null hypothesis (all trials, $r_{14} = .06$, $p = .8$, BF = 5.2; grapheme trials, $r_{14} = -.2$, $p = .5$, BF = 4.1; color trials, $r_{14} = .1$, $p = .6$, BF = 4.6). Again, the results were similar when those who only completed the SB when it came back online were excluded (all trials, $r_{12} = .06$, $p = .8$, BF = 4.9; grapheme trials, $r_{12} = -.1$, $p = .7$, BF = 4.6; color trials, $r_{12} = .1$, $p = .7$, BF = 4.6). There was also no significant correlation between congruency magnitudes for accuracy and mean Euclidean distance in RGB space, averaged across the two graphemes, overall ($r_{15} = -.2$, $p = .4$, BF = 3.6), or for either grapheme ($r_{15} = -.3$, $p = .3$, BF = 3.1) or color ($r_{15} = -.2$, $p = .5$, BF = 4.5: Figure 6B) trials (Bonferroni-corrected α for three tests = .017); this was also the case for correlations with the mean Euclidean distances calculated for each grapheme separately.

3.2.3 Within-test variability—Since the SB score and Euclidean distances in RGB color space reflect within-test variability in the specific colors picked to correspond with graphemes, we also compared these to the within-test variability of the IAT congruency magnitudes. We calculated the congruency magnitudes separately for each of the two runs of the IAT and subtracted those for the first run from those for the second run to arrive at the within-test change in congruency magnitudes, for all trials combined and separately for grapheme and color trials. However, there were no significant correlations between these measures and SB scores (Bonferroni-corrected α for three tests = .017) for either RTs ($r_{15} = .2 - .3$, $p = .2 - .4$, BFs = 2.8 – 3.6: Figure 7A) or accuracy ($r_{15} = .3 - .4$, $p = .1 - .2$, BFs = 1.5 – 2.6: Figure 7B). The RT and accuracy data each contained a single outlier, but excluding these data points did not change the significance of the results in either case, and if anything tended to make the BFs larger (RTs, $r_{14} = .1 - .2$, $p = .5 - .6$, BFs = 4.2 – 4.7; accuracy, $r_{14} = -.2 - -.4$, $p = .1 - .5$, BFs = 1.7 – 4.2). Again, similar results were obtained when those who only completed the SB when it came back online were excluded (RTs, $r_{12} = .3 - .4$, $p = .2 - .3$, BFs = 1.9 – 2.7; accuracy, $r_{12} = -.2 - -.3$, $p = .2 - .5$, BFs = 2.5 – 4.5). Similarly, there were no significant correlations between the congruency magnitude difference between runs and mean Euclidean distance in RGB color averaged across the pair of graphemes (Bonferroni-corrected α for three tests = .017) for either accuracy (all $r_{15} = .3$, all $p = .2$, BF = 2.7 – 2.9) or RTs ($r_{15} = .2 - .3$, $p = .2 - .3$, BF = 2.1 – 3.5: Figure 8). Here too, excluding the outliers mentioned above did not affect the significance of the results (RTs, $r_{14} = .3 - .4$, $p = .1 - .3$, BFs = 1.6 – 3.0; accuracy, $r_{14} = -.1 - -.2$, $p = .5 - .7$, BFs = 4.1 – 5.0). Again, there were still no significant consistency-strength relationships when the correlations were run against the mean Euclidean distances calculated for each grapheme separately.

3.2.4 Additional supporting analyses—As noted above, Euclidean distances were derived from RGB values extracted from either SB data or the online color-picker which were completed prior to performing the IAT. Since these RGB values were all obtained close to the time of testing, we consider these to be the most reliable. For completeness, we also re-ran the correlations, replacing Euclidean distances derived from the online color-picker with those derived from SB data when it came back online for the three synesthetes who did so. However, this did not change the results: there were still no significant relationships between strength and consistency when the latter was based on

Euclidean distance, including the within-test variability measures (all $r_{15} = -.3 - .3$, all p = .1 – .9, all BFs = 1.9 – 5.4).

Finally, non-significant correlations might occur because of a lack of variation in the data across participants. Table 1 provides measures of variability, with and without outliers, in the consistency (SB scores and mean Euclidean distance), strength (IAT congruency magnitude), and the within-test change in congruency magnitude data, allowing a quantitative judgment that the lack of significant correlations was not due to a lack of variance in the data. The scatterplots and trend lines in Figures 5–8 similarly allow a qualitative judgment that, despite this variance, there was no significant relationship between consistency and strength. The presence of outliers might suggest that the variation across subjects is over-stated. However, in the current data, all correlations were non-significant whether the outliers were excluded or not, even before correcting the alpha-value for multiple tests and despite the multiple ways in which the data were examined.

4. DISCUSSION

The present study is the first, to our knowledge, to assess the strength of synesthetic associations using the IAT and to compare strength to consistency. Congruency magnitudes on the IAT distinguished synesthetes from non-synesthetes, with a large effect size especially when RT was the dependent variable (d = 2.0), based on which we propose that the strength of synesthetic associations is a defining characteristic of synesthetic experience (though not, of course, decisive on its own). If confirmed in future studies, this would be a novel, objective marker of synesthesia. Intuitively, we expected that strong associations should be consistent and vice versa, and that therefore strength and consistency measures would be correlated. Surprisingly, this was not the case. IAT congruency magnitudes were uncorrelated with consistency measures, whether the SB score, reflecting overall consistency of the colors evoked by a large grapheme set for which an individual has associations, or the mean Euclidean distance in RGB color space for the two graphemes tested for each synesthete with the IAT, reflecting consistency for these specific grapheme-color associations. Even the within-test variability of our strength measure, the IAT congruency magnitude, was uncorrelated with the consistency measures, which reflect within-test variability of the color selected to match a particular grapheme. Importantly, the BFs associated with the correlational analyses were generally greater than 3, indicating substantial support for accepting the null hypothesis (Jeffreys, 1961; Dienes, 2014). This was particularly the case for analyses based on RTs, perhaps reflecting that accuracy is a less sensitive measure due to the potential for ceiling effects. Overall, these results suggest that the strength and consistency of associations are separable aspects of synesthetic experience.

The consistency of synesthetic associations may remain stable over long periods of time (Asher et al., 2006; Linkovski et al., 2012; Simner & Logie, 2007) but there can also be long-term changes due to age (Meier et al., 2014; Simner et al., 2017; Chromý et al., 2019) and temporary changes due to mood (Kay et al., 2015). Here, we only measured consistency over the short time it took to complete the SB or the online color-picker task; variability of strength was also assessed over a similar time scale. Further evidence is required from long-term studies that assess strength and consistency at multiple time points

in order to determine whether the strength of synesthetic association is stable over time even if the particular concurrent has changed, leading to low consistency. Also, the strength of synesthetic associations was examined in the present study in relation to two graphemes evoking very distinct colors. It would be interesting to assess whether the strength measure is stable across a wider range of synesthetic associations, including those in types of synesthesia other than grapheme-color.

If synesthetic associations can be strong but variable, this has implications for identification of synesthesia and estimates of its prevalence. For example, SB scores below 1 denote the presence of synesthesia whereas non-synesthetes demonstrate SB scores above 2, with a score falling between 1 and 2 being taken to mean that synesthetic status cannot be reliably determined (Eagleman et al., 2007). Scores in this indeterminate range may be obtained for a number of potential reasons, e.g. people falsely claiming to be synesthetic, misunderstandings of the nature of synesthesia, or potential dissociation between synesthetic strength and consistency in synesthetes whose inducers strongly evoke a concurrent but not one that is stable enough to pass the consistency test (Simner, 2012). One might argue that an individual still needs at least two consistent inducer-concurrent mappings in order to complete the IAT. But the SB, and other tests of consistency, are predicated on synesthetes reporting exactly the same color for any given grapheme whereas the IAT need not depend on this. For example, 'A' might have a strong association with red, but the exact shade might not be consistent trial by trial. For 'indeterminate' synesthetes, it will be important to take a careful history of their grapheme-color associations and how these are experienced, for effective testing with the IAT; converging evidence, e.g., from the Stroop test (Mattingley et al., 2001; Dixon et al., 2000, 2004; Linkovski et al., 2012; Akiva-Kabiri et al., 2014) may also be required. Additionally, synesthesia may be a phenomenon at one end of a continuum with no clear cut-off point on any characteristic (Eagleman, 2012), thus there may also be low strength/high consistency synesthetes who pass the SB threshold that could be identified by such studies. Future studies, combining tests of consistency with tests of strength and including individuals across the full range of SB scores might help to disambiguate high strength/low consistency synesthetes from non-synesthetes in the SB's indeterminate category, thus increasing estimates of prevalence. While it is hard to assess the scale of any under-detection of synesthesia, it is worth noting that of the eight people who volunteered for this study claiming to experience grapheme-color synesthesia but who were excluded on the basis of their SB score, only one failed the consistency test outright while the remaining seven, who scored in the indeterminate range, may have included genuine synesthetes (see Footnote 1).

Both the IAT (Greenwald et al., 1998) and the classic Stroop (1935) test have been used to investigate aspects of synesthesia; both rely on differences in RT/accuracy between congruent and incongruent trials for their effects. As the present study is only the second, to our knowledge, to use the IAT in the context of synesthesia, it is worth comparing its use to that of the more frequently used Stroop test. Despite its name, the IAT captures both implicit and explicit associations (Blair, 2002; Fiedler et al., 2006): a trial is characterized as congruent when related concepts, e.g., 'cold' and 'ice' are associated with the same response key and incongruent when the same response key is used for unrelated or opposing concepts, e.g., 'cold' and 'steam'. IAT effects thus arise at the response level from stimulus-response

(in)congruity (De Houwer, 2001). This is borne out in the present study: congruency magnitudes for color trials were more affected by the (in)congruity of the response key associations than the perceptual (in)congruity of the synesthetic color evoked during the preceding grapheme trial. Stroop effects, on the other hand, may arise at least in part at the perceptual level from stimulus-stimulus (in)congruity (Scerrati et al., 2017), i.e., for the word "blue" displayed in red, it is harder to name the print-color than the word. However, the underlying basis of the Stroop effect remains controversial (Parris et al., 2019), with varying accounts including response-related processes such as conflict and cognitive control (Botvinick et al., 2001) versus earlier processes like selective attention (McLeod, 1991, 1992; Algom and Chajut, 2019), and neuroimaging studies suggesting multiple loci of the effect (Banich, 2019). For grapheme-color synesthesia, a trial in the Stroop test is considered congruent when a letter is displayed in its associated synesthetic color and incongruent if displayed in any other color (e.g., Mattingley et al., 2001; Dixon et al., 2004).

The IAT and the synesthetic Stroop test both likely reflect both the strength of synesthetic associations and the automaticity of the synesthetic association (Mattingley et al., 2001; Dixon et al., 2004) but are methodologically very different. One advantage of the Stroop test is that (in)congruity is inherent in the stimulus and therefore congruent and incongruent trials can be randomly ordered (though this need not be so – see Mattingley et al., 2001). By contrast, congruent and incongruent trials for the IAT have to be presented in blocks because changing the response key associations on a trial-by-trial basis would be too onerous. On the other hand, the IAT presents only one stimulus at a time and thus avoids the confound of divided attention (Parise & Spence, 2012) whereas attention is divided between the word and its print-color in the Stroop test (McLeod, 1991). Additionally, whereas the IAT can only test two graphemes at a time, the synesthetic Stroop could efficiently present all graphemes for which a particular synesthete has color associations. Further work is needed on the relationship between the synesthetic versions of the Stroop and IAT, and to explore whether synesthetic Stroop effects are also uncorrelated with consistency.

In both the present study, directly addressing synesthetic experience, and our previous study of synesthetic responses to crossmodal correspondences (Lacey et al., 2016), the IAT reliably distinguished between synesthetes and non-synesthetes, as has the Stroop test (e.g., Mattingley et al., 2001; Dixon et al., 2000, 2004; Linkovski et al., 2012; Akiva-Kabiri et al., 2014) and thus the IAT appears to be a useful addition to the field. Beyond grapheme-color, the Stroop test has also authenticated different kinds of synesthesia, e.g., pitch class-color (Itoh et al., 2019), pitch-space (Linkovski et al., 2012; Akiva-Kabiri et al., 2014), sequence-space (Ward et al., 2018), and even swimming style-color (Nikolić et al., 2011). In grapheme-color synesthesia, the Stroop test also distinguishes between projector and associator synesthetes (Dixon et al., 2004). A useful goal for future research will be to extend the use of the IAT to other synesthesias – given that it presents one stimulus at a time, it may be particularly useful for inducer-concurrent pairings that are less amenable to the Stroop paradigm, e.g., lexical-gustatory or ordinal-personality synesthesias – and to see whether it is also sensitive to the projector-associator dimension.

The neural mechanisms underpinning the strength of synesthetic associations are uncertain. A functional magnetic resonance imaging (fMRI) study of grapheme-color synesthetes

performing the synesthetic Stroop test showed that, compared to non-synesthetic control participants, synesthetes had stronger anterior cingulate responses to graphemes displayed in an incongruent color (van der Veen et al., 2014). This is consistent with the role of the cingulate cortex in attention (e.g., Vogt, 2019) and conflict detection and resolution (Li et al., 2017). Banich (2019) suggests that the anterior cingulate is only involved at the last stage of a processing 'cascade' originating in posterior lateral prefrontal cortex, directing processing resources based on task-relevance of information, and proceeding via mid-dorsolateral and mid-cingulate regions that bias working memory towards task-relevant information and response selection (Banich, 2019). There are no fMRI studies in which the strength of synesthetic associations is tested using the IAT. However, the large effect of synesthetic congruency particularly on RT (d = 2.0), which was greater than on accuracy (d = .8) suggests that decision and response-related neural processes are likely to be important. This is in keeping with the idea that IAT effects arise at the response level (de Houwer, 2001) and our finding that (in)congruity of response key associations had a greater effect on color trials than (in)congruity of the synesthetic color evoked during the preceding grapheme trial. By contrast, consistency measures may reflect perceptual processes to a greater extent. This may explain why consistency and strength measures were uncorrelated. Nonetheless, it is also possible that a more perceptually-based measure of strength could demonstrate different properties, and perhaps correlate with consistency measures. It is interesting that performance differences on the IAT between synesthetes and controls were much greater on incongruent than congruent trials; this was especially true for RT. This suggests that synesthetes have particular difficulty with inhibiting their synesthetic associations, and, relative to this incongruency cost, derive a smaller benefit from congruent associations. Thus, we propose that the neural mechanisms involved in synesthesia as indexed by the IAT might depend on frontoparietal processes that mediate perceptual decisions and/or responses. Further work is necessary to test this proposal.

Consistency and strength measures of synesthesia may reflect different underlying neural mechanisms, which will need to be isolated from the task-related effects just described. Dovern et al. (2012), using the fMRI signal to measure resting-state functional connectivity (RSFC) between networks identified by independent component analysis, found that grapheme-color synesthetes exhibited higher RSFC for some inter-network connections compared to controls.

Among these connections, those connecting a visual network including bilateral fusiform gyri to an auditory network that included both superior temporal gyri and to a right frontoparietal network displayed RSFC that correlated with the synesthetes' SB scores; notably, these SB scores were confirmed to be stable over a 6-month period. The fusiform gyri are implicated in processing both color (Shapley and Hawken, 2011) and graphemes (Pernet et al., 2005), while frontoparietal regions may be involved in higher-order post-perceptual processes related to synesthesia (Ward et al., 2006; Brang et al., 2011; Chiou & Rich, 2014). The magnitude of RSFC appears to reflect the robustness of bi-directional connections in macaque monkeys (Pijnenburg et al., 2019). Thus, it is interesting that the magnitude of RSFC between networks comprising the fusiform and frontoparietal cortex correlates with the consistency of synesthetic association, as indexed by the SB score (Dovern et al., 2012). How precisely RSFC magnitude comes to be correlated with

synesthetic consistency, and whether RSFC magnitude correlations with synesthetic strength would reveal different networks, remain to be determined.

5. CONCLUSIONS

Our findings suggest that consistency and strength of synesthetic associations are separable aspects of synesthetic experience and that strength is a novel marker of synesthesia. Longitudinal studies are required to assess whether the strength of association for a particular grapheme remains stable over time despite variability in the associated color. This could help to resolve the synesthetic status of those who fall within the indeterminate range of SB scores. Strength and consistency may depend on distinct neural mechanisms. We suggest that, while the consistency of grapheme-color synesthesia may depend primarily on perceptual processes that are related to the magnitude of connectivity between the fusiform gyrus and frontoparietal cortex, the strength of the relevant associations as indexed by the IAT is likely to depend more on frontoparietal processing related to decision and response. A quest for more perceptually based measures of synesthetic strength merits pursuit.

ACKNOWLEDGEMENTS

REFERENCES

Akiva-Kabiri L, Linkovski O, Gertner L. & Henik A. (2014). Musical-space synesthesia: automatic, explicit and conceptual connections between musical stimuli and space. Consciousness & Cognition, 28:17–29. [PubMed: 25023138]

Anikin A. & Johansson N. (2019). Implicit associations between individual properties of color and sound. Attention, Perception, & Psychophysics, 81:764–777.

Asher JE, Aitken MR, Farooqi N, Kurmani S. & Baron-Cohen S. (2006). Diagnosing and phenotyping visual synesthesia: a preliminary evaluation of the revised test of genuineness (TOG-R). Cortex, 42:137–146. [PubMed: 16683486]

Banich MT (2019). The Stroop effect occurs at multiple points along a cascade of control: evidence from cognitive neuroscience approaches. Frontiers in Psychology, 10:2164, doi:10.3389/fpsyg.2019.02164 [PubMed: 31681058]

Blair IV (2002). The malleability of automatic stereotypes and prejudice. Personality & Social Psychology Review, 6:242–261.

Botvinick MM, Braver TS, Barch DM, Carter CS & Cohen JD (2001). Conflict monitoring and cognitive control, Psychological Review, 108:624–652. [PubMed: 11488380]

Brang D, Rouw R, Ramachandran VS & Coulson S. (2011). Similarly shaped letters evoke similar colors in grapheme-color synesthesia. Neuropsychologia, 49:1355–1358. [PubMed: 21219918]

Carmichael DA, Down MP, Shillock RC, Eagleman DM & Simner J. (2015). Validating a standardised test battery for synesthesia: does the Synesthesia Battery reliably detect synesthesia? Consciousness & Cognition, 33:375–385. [PubMed: 25734257]

Chan ZPY & Dyson BJ (2015). The effects of association strength and cross-modal correspondence on the development of multimodal stimuli. Attention, Perception, & Psychophysics, 77:560–570.

Chiou R. & Rich AN (2014). The role of conceptual knowledge in understanding synesthesia: evaluating contemporary findings from a 'hub-and-spokes' perspective. Frontiers in Psychology, 5:105, doi: 10.3389/fpsyg.2014.00105 [PubMed: 24653707]

Chromý J, Bor vková M, Malá L. & Sudzinová T. (2019). Long-term versus short-term consistency in the grapheme-colour synesthesia: grapheme-colour pairings can change in adulthood. Attention, Perception, & Psychophysics, 81:1806–1812.

Cumming G. (2014). The new statistics: why and how. Psychological Science, 25:7–29. [PubMed: 24220629]

De Houwer J. (2001). A structural and process analysis of the Implicit Association Test. Journal of Experimental Social Psychology, 37:443–451.

Dienes Z. (2014). Using Bayes to get the most out of non-significant results. Frontiers in Psychology, 5:781, doi:10.3389/fpsyg.2014.00781 [PubMed: 25120503]

Dixon MJ & Smilek D. (2005). The importance of individual differences in grapheme-color synesthesia. Neuron, 45:821–823. [PubMed: 15797543]

Dixon MJ, Smilek D, Cudahy C. & Merikle PM (2000). Five plus two equals yellow. Nature, 406:365. [PubMed: 10935623]

Dixon MJ, Smilek D, & Merikle PM (2004). Not all synaesthetes are created equal: Projector versus associator synaesthetes. Cognitive, Affective, & Behavioral Neuroscience, 4:335–343.

Dovern A, Fink GR, Fromme CB, Wohlschläger AM, Weiss PH et al. (2012). Intrinsic network connectivity reflects consistency of synesthetic experiences. Journal of Neuroscience, 32:7614–7621. [PubMed: 22649240]

Eagleman DM (2012). Synaesthesia in its protean guises. British Journal of Psychology, 103:16–19 [PubMed: 22229769]

Eagleman DM, Kagan AD, Nelson SS, Sagaram D. & Sarma AK (2007). A standardized test battery for the study of synesthesia. Journal of Neuroscience Methods, 159:139–145. [PubMed: 16919755]

Fiedler K, Messner C. & Bluemke M. (2006). Unresolved problems with the "I", the "A", and the "T": a logical and psychometric critique of the implicit association test (IAT). European Review of Social Psychology, 17:74–147.

Gebuis T, Nijboer TCW & Van der Smagt MJ (2009). Multiple dimensions in bi-directional synesthesia. European Journal of Neuroscience, 29:1703–1710. [PubMed: 19419432]

Gebuis T, Nijboer TC & van der Smagt MJ (2009a). Of colored numbers and numbers colors: interactive processes in grapheme-color synesthesia. Experimental Psychology, 56:180–187. [PubMed: 19289360]

Greenwald AG, McGhee DE & Schwarz JLK (1998). Measuring individual differences in implicit cognition: the implicit association test. Journal of Personality & Social Psychology, 74:1464–1480. [PubMed: 9654756]

Hoijtink H, van Kooten P. & Hulsker K. (2016). Why Bayesian psychologists should change the way the use the Bayes factor. Multivariate Behavioral Research, 51:2–10. [PubMed: 26881951]

Hubbard EM, Manohar S. & Ramachandran VS (2006). Contrast affects the strength of synesthetic colors. Cortex, 42:184–194. [PubMed: 16683492]

Hunt RWG (2004). The Reproduction of Colour (6th edition). Wiley: Chichester, UK.

Ipser A, Ward J. & Simner J. (2020). The MULTISENSE test of lexical-gustatory synesthesia: an automated online diagnostic. Behavior Research Methods, 52:544–560. [PubMed: 31161427]

Itoh K, Sakata H, Igarashi H. & Nakada T. (2019). Automaticity of pitch class-color synesthesia as revealed by a Stroop-like effect. Consciousness & Cognition, 71:86–91. [PubMed: 30978617]

Jarosz AF & Wiley J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. Journal of Problem Solving, 7, doi:10.7771/1932-6246.1167

Jeffreys H. (1961). Theory of Probability (3rd edition). Oxford, UK: Oxford University Press.

Kay CL, Carmichael DA, Ruffell HE & Simner J. (2015). Colour fluctuations in grapheme-colour synesthesia: the effect of clinical and non-clinical mood changes. British Journal of Psychology, 106:487–504. [PubMed: 25413977]

Kriegekorte N, Mur M. & Bandettini P. (2008b). Presentational similarity analysis – connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2:4, doi:10.3389/neuro.06.004.2008 [PubMed: 19104670]

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H. et al. (2008a). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60, 1126–1141. [PubMed: 19109916]

Lacey S, Martinez M, McCormick K. & Sathian K. (2016). Synesthesia strengthens sound-symbolic cross-modal correspondences. European Journal of Neuroscience, 44:2716–2721. [PubMed: 27564319]

Lenhard W. & Lenhard A. (2016). Calculation of Effect Sizes. Psychometrica. doi: 10.13140/RG.2.1.3478.4245 Accessed at https://www.psychometrica.de/effect_size.html.

Li Q, Yang G, Li Z, Qi Y, Cole MW & Liu X. (2017). Conflict detection and resolution rely on a combination of common and distinct cognitive control networks. Neuroscience & Biobehavioral Reviews, 83:123–131. [PubMed: 29017916]

Linkovski O, Akiva-Kabiri L, Gertner L. & Henik A. (2012). Is it for real? Evaluating authenticity of musical pitch-space synesthesia. Cognitive Processing, 13 (Supplement 1):S247–S251. [PubMed: 22806677]

Makin ADJ & Wuerger SM (2013). The IAT shows no evidence for Kandinsky's color-shape associations. Frontiers in Psychology, 4:616, doi:10.3389/fpsyg.2013.00616 [PubMed: 24062709]

Mattingley JB, Rich AN, Yelland G. & Bradshaw JL (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. Nature, 410:580–582. [PubMed: 11279495]

McLeod CM (1992). The Stroop task: The "gold standard" of attentional measures. Journal of Experimental Psychology: General, 121:12–14.

McLeod CM (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109:163–203. [PubMed: 2034749]

Meier B, Rothen N. & Walter S. (2014). Developmental aspects of synaesthesia across the adult lifespan. Frontiers in Human Neuroscience, 8:129 doi:10.3389/fnhum.2014.00129 [PubMed: 24653689]

Menouti K, Akiva-Kabiri L, Banissy MJ & Stewart L. (2015). Timbre-colour synaesthesia: exploring the consistency of associations based on timbre. Cortex, 63:1–3. [PubMed: 25241393]

Niccolai V, Jennes J, Stoerig P. & van Leeeuwen TM (2012). Modality and variability of synesthetic experience. American Journal of Psychology, 125:81–94. [PubMed: 22428428]

Nikoli D, Jürgens UM, Rothen N, Meier B. & Mroczko A. (2011). Swimming-style synesthesia. Cortex, 47:874–879. [PubMed: 21402377]

Novich S, Cheng S. & Eagleman DM (2011). Is synaestheisa one condition or many? A large-scale analysis reveals subgroups. Journal of Neuropsychology, 5:353–371. [PubMed: 21923794]

Ovalle Fresa R. & Rothen N. (2019). Development of synaesthetic consistency: repeated autonomous engagement with graphemes and colours leads to consistent associations. Consciousness & Cognition, 73:102764.

Parise CV & Spence C. (2012). Audiovisual cross-modal correspondences and sound symbolism: a study using the implicit association test. Experimental Brain Research, 220:319–333. [PubMed: 22706551]

Pernet C, Celsis P. & Demonet J. (2005). Selective response to letter categorization within the left fusiform gyrus. NeuroImage 28:738–744. [PubMed: 16109492]

Pijnenburg R, Scholtens LH, Mantini D, Vanduffel W, Feldman Barrett L. et al. (2019). Biological Characteristics of Connection-Wise Resting-State Functional Connectivity Strength. Cerebral Cortex, 29:4646–4653. [PubMed: 30668705]

Rothen N, Jünemann K, Mealor AD, Burckhardt V. & Ward J. (2016). The sensitivity and specificity of a diagnostic test of sequence-space synesthesia. Behavior Research Methods, 48:1476–1481. [PubMed: 26558902]

Rothen N, Seth AK, Witzel C. & Ward J. (2013). Diagnosing synesthesia with online colour pickers: maximizing sensitivity and specificity. Journal of Neuroscience Methods, 215:156–160. [PubMed: 23458658]

Scerrati E, Lugli L, Nicoletti R. & Umiltà C. (2017). Comparing Stroop-like and Simon effects on perceptual features. Scientific Reports, 7:17815, doi: 10.1038/s41598-017-18185-1 [PubMed: 29259275]

Shalgi S. & Foxe JJ (2009). The neurophysiology of bi-directional synesthesia (Commentary on Gebuis et al.) European Journal of Neuroscience, 29:1701–1702. [PubMed: 19419431]

Shapley R. & Hawken MJ (2011). Color in the Cortex: single- and double-opponent cells. Vision Research, 51:701–717. [PubMed: 21333672]

Simner J. (2012). Defining synaesthesia. British Journal of Psychology, 103:1–15. [PubMed: 22229768]

Simner J. & Bain AE (2013). A longitudinal study of grapheme-color synesthesia in childhood: 6/7 years to 10/11 years. Frontiers in Human Neuroscience, 7:603 doi:10.3389/fnhum.2013.00603 [PubMed: 24312035]

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. [PubMed: 28919244]

Simner J. & Logie RH (2007). Synaesthetic consistency spans decades in a lexical-gustatory synaesthete. Neurocase, 13:358–365. [PubMed: 18781434]

Stroop JR (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18:643–662.

Uno K, Asano M. & Yokosawa K. (2021). Consistency of synesthetic association varies with grapheme familiarity: A longitudinal study of grapheme-color synesthesia. Consciousness & Cognition, 89:103090

Van der Veen FM, Aben HP, Smits M. & Röder CH (2014). Grapheme-color synesthesia interferes with color perception in a standard Stroop task. Neuroscience, 258:246–253. [PubMed: 24269938]

Vogt BA (2019). Cingulate impairments in ADHD: Comorbidities, connections, and treatment. In Vogt BA (Ed.) Handbook of Clinical Neurology, vol. 166, pp297–314: Elsevier BV, Netherlands. [PubMed: 31731917]

Ward J. (2013). Synesthesia. Annual Review of Psychology, 64:49–75.

Ward J, Huckstep B. & Tsakanikos E. (2006) Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? Cortex, 42:264–280. [PubMed: 16683501]

Ward J, Ipser A, Phanvanova E, Brown P, Bunte I. & Simner J. (2018). The prevalence and cognitive profile of sequence-space synesthesia. Consciousness & Cognition, 61:79–93. [PubMed: 29673773]

Weiss PH, Kalckert A. & Fink GR (2008). Priming letters by colors: evidence for the bi-directionality of grapheme-color synesthesia. Journal of Cognitive Neuroscience, 21:2019–2026.

Zellner A. & Siow A. (1980). Posterior odds ratios for selected regression hypotheses. In Bernardo JM, De Groot MH, Lindley DV & Smith AFM (eds.), Bayesian Statistics, pp585– 603. Valencia: University Press.

HIGHLIGHTS

- Synesthetic status is usually determined by testing consistency of associations

- Consistency may not be entirely reliable: associations can change with age or mood

- We compared consistency to strength of association in grapheme-color synesthetes

- Strength identified synesthetes but was unrelated to consistency

- Synesthetes may be under-detected by tests that rely on consistency
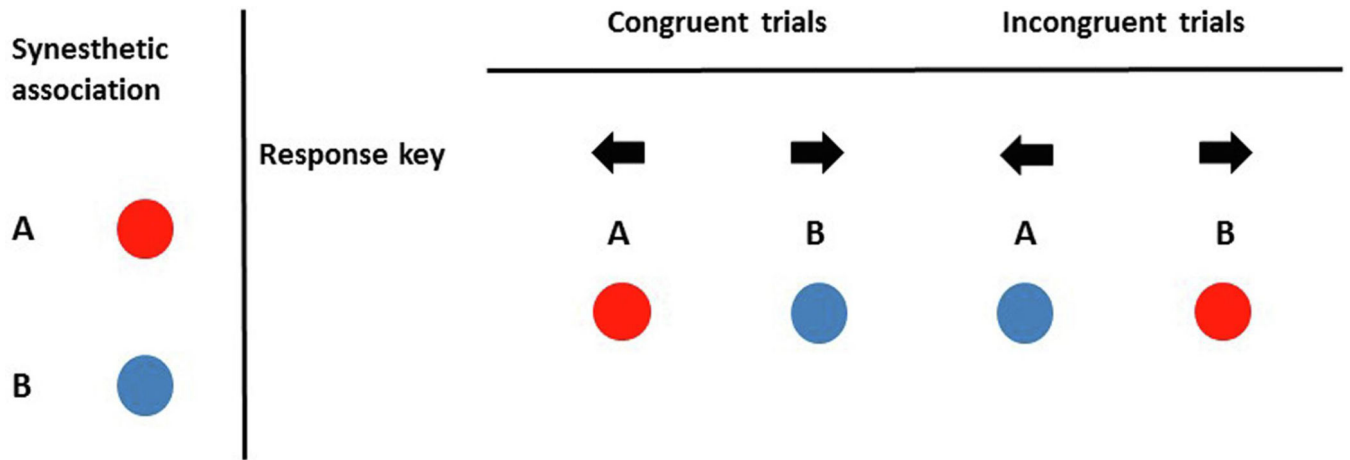
Figure 1:
Schematic of example synesthetic associations and response key pairings used in the IAT experiment.
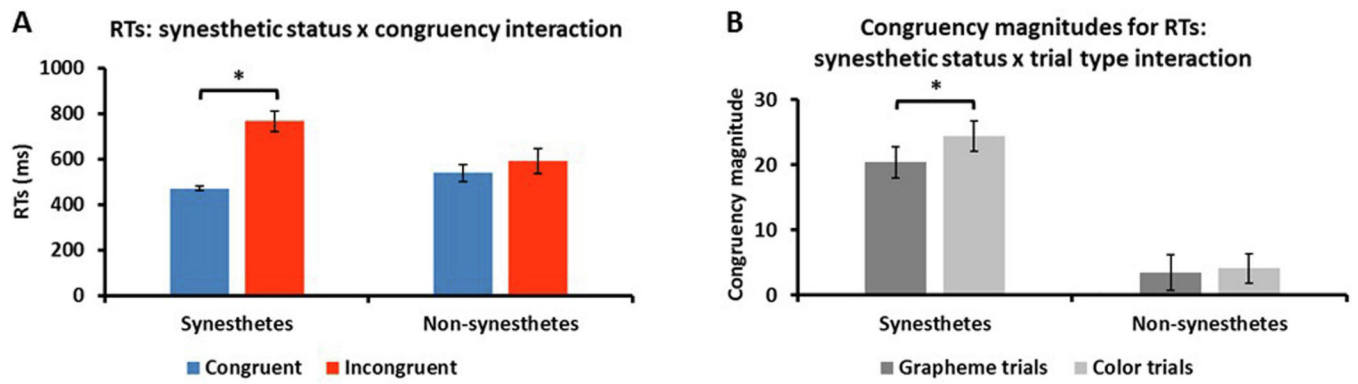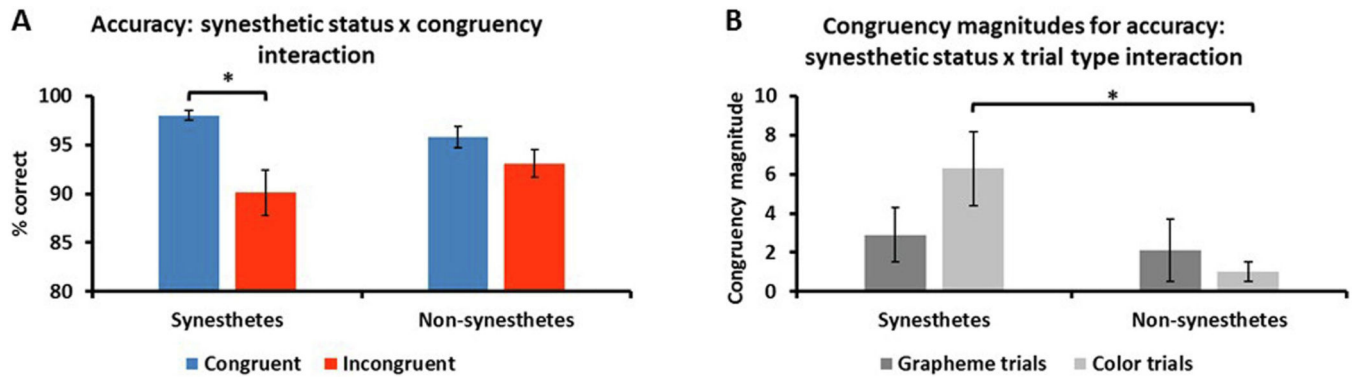
Figure 2:
Interaction between (A) synesthetic status and congruency for RTs (correct trials only) in which synesthetes were significantly faster to respond on congruent, compared to incongruent, trials whereas there was no significant difference for non-synesthetes. * p < .001; (B) synesthetic status and trial type for RT congruency magnitudes in which synesthetes exhibited significantly larger congruency magnitudes on color, compared to grapheme, trials whereas there was no significant difference for non-synesthetes. * p = .007; error bars = sem.

GC trials: synesthetic status x magnitude type interaction
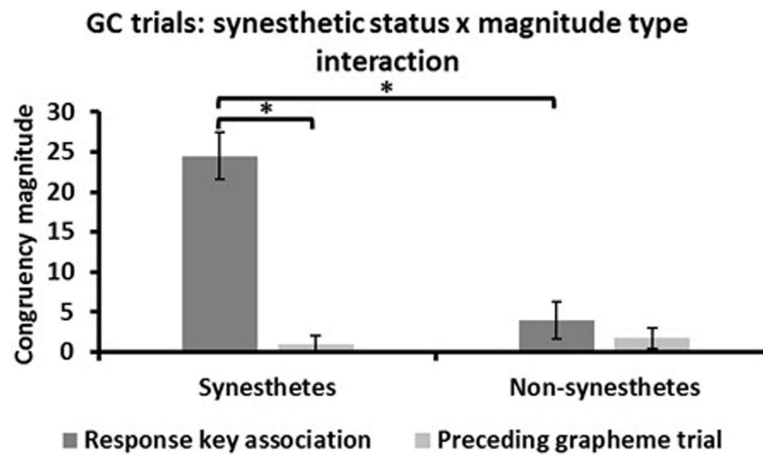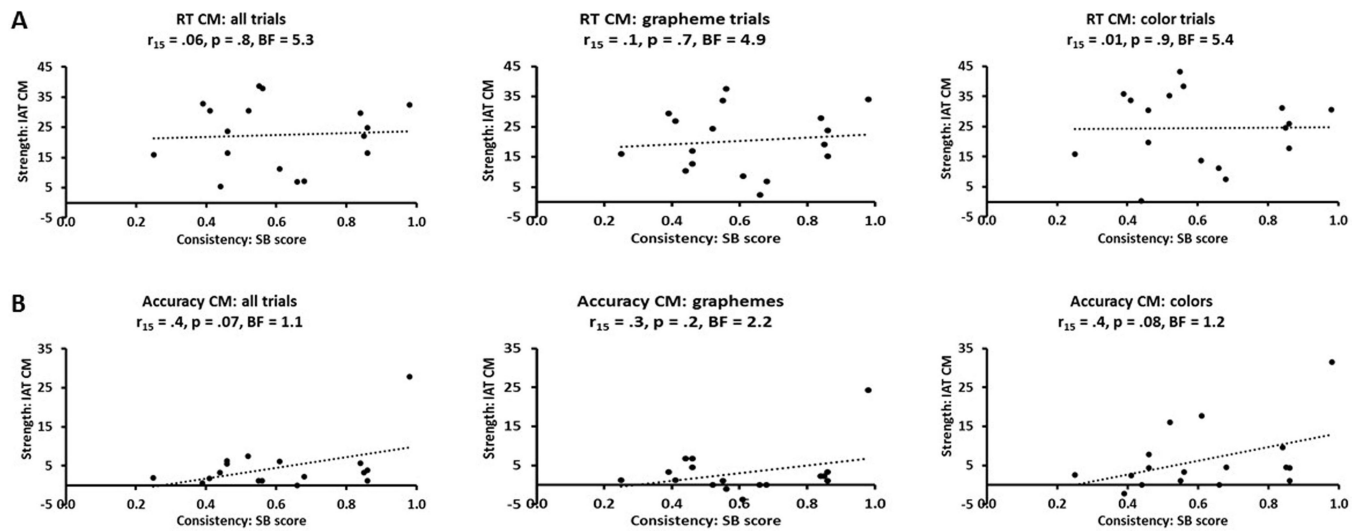
Figure 3:

Analysis of RT congruency magnitudes for color trials (correct trials only) that were immediately preceded by grapheme trials (GC trials) showed that synesthetes were more affected by response key associations than the preceding grapheme trial while non-synesthetes were unaffected by either. Additionally, synesthetes were more affected by response key associations than non-synesthetes but neither were affected by the preceding trial. * p < .001; error bars = sem.
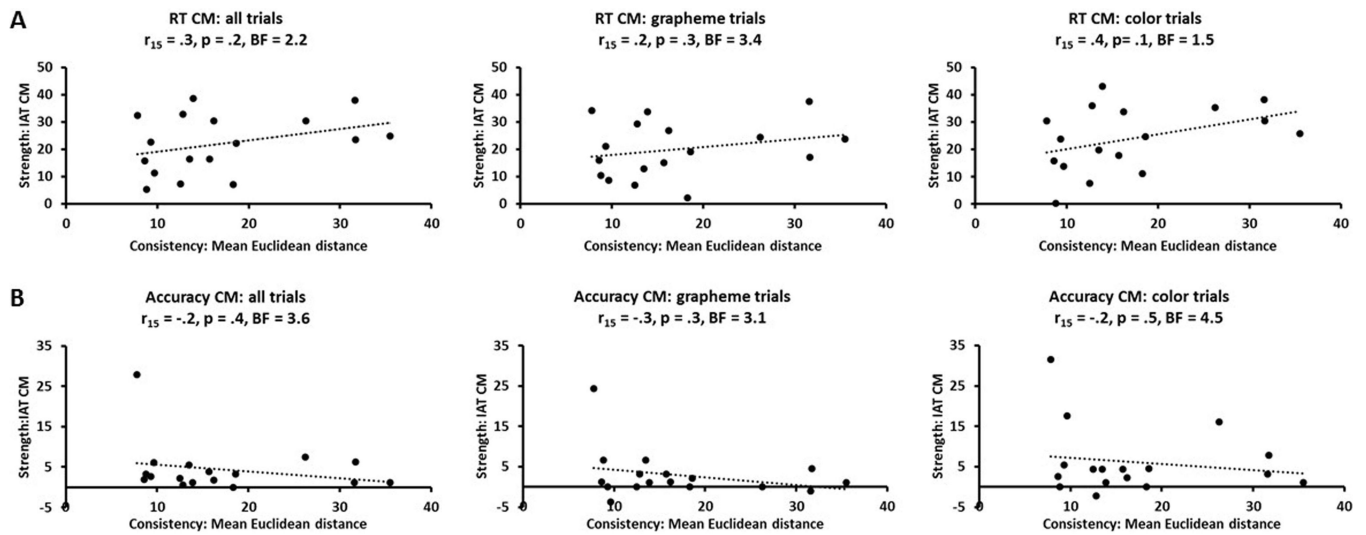
Figure 4:
Interaction between (A) synesthetic status and congruency for accuracy in which synesthetes were significantly more accurate on congruent, compared to incongruent, trials whereas there was no significant difference for non-synesthetes. * p = .003; (B) synesthetic status and trial type for accuracy congruency magnitudes in which synesthetes exhibited significantly larger congruency magnitudes than non-synesthetes on color, but not grapheme, trials; there was no significant difference between grapheme and color magnitudes within each group. * p = .01; error bars = sem.
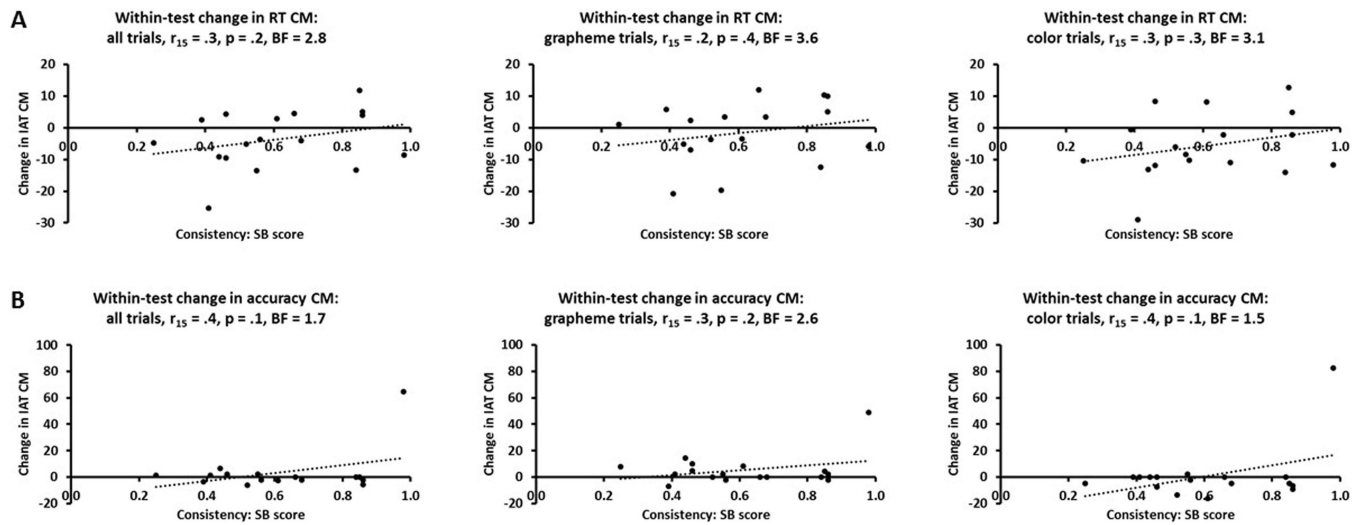
Figure 5:
Scatterplots showing that consistency (SB score) and strength (IAT congruency magnitudes) of synesthetic grapheme-color associations were uncorrelated for both (A) RT and (B) accuracy, whether these were calculated across all trials or for grapheme and color trials separately. CM = congruency magnitude; BF = Bayes factor.
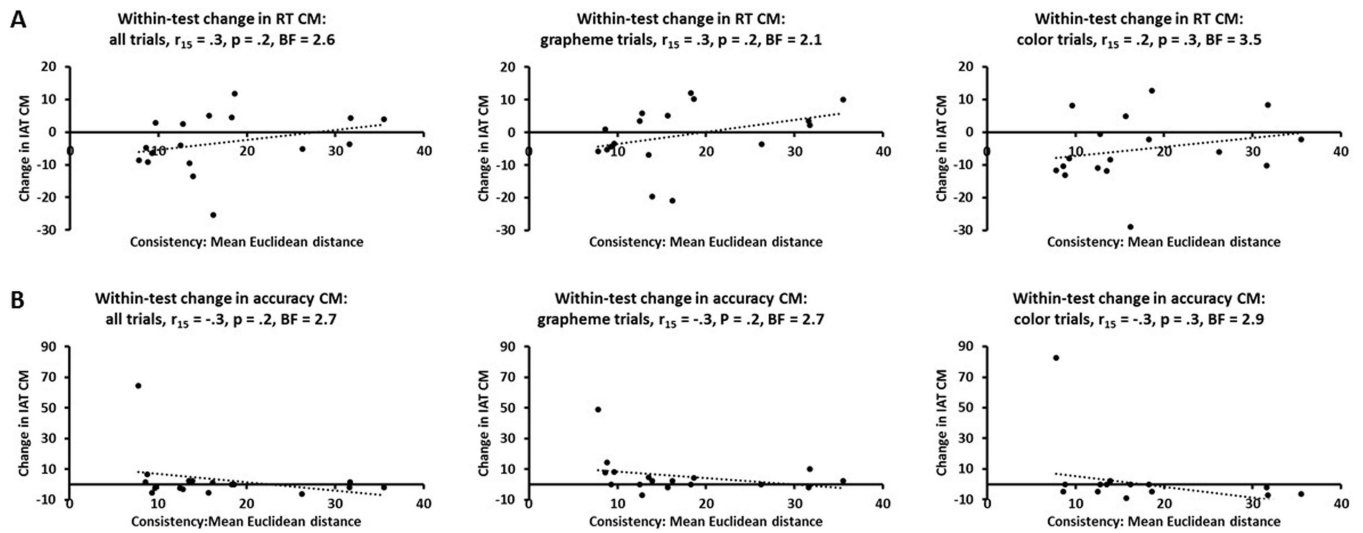
Figure 6:
Scatterplots showing that consistency (mean Euclidean distance in RGB space for the three color estimates for the two graphemes) and strength (IAT congruency magnitudes) of synesthetic grapheme-color associations were uncorrelated for both (A) RT and (B) accuracy, whether these were calculated across all trials or for grapheme and color trials separately. CM, BF as for Figure 5.

Figure 7.
Scatterplots showing that the consistency (SB score) of synesthetic grapheme-color associations and the within-test change in IAT congruency magnitudes were uncorrelated for both (A) RT and (B) accuracy, whether these were calculated across all trials or for grapheme and color trials separately. CM, BF as for Figure 5.

Figure 8.
Scatterplots showing that consistency (mean Euclidean distance in RGB space for the three color estimates for the two graphemes) of synesthetic grapheme-color associations and the within-test change in IAT congruency magnitudes were uncorrelated for both (A) RT and (B) accuracy, whether these were calculated across all trials or for grapheme and color trials separately. CM, BF as for Figure 5.

TABLE 1:

Variability in measures of (A) consistency, (B) strength of synesthetic associations as indexed by IAT congruency magnitudes, and (C) within-test change in IAT magnitudes: [excluding outliers].

		Mean	Standard deviation	Range
(A) Synesthesia Battery score		.61	.21	0.73
Mean Euclidean distance		17.1	8.9	27.7
(B) RT congruency magnitudes	Total	22.5 [22.2]	10.6 [10.7]	33.2 [-]
	Grapheme	20.4 [20.0]	10.1 [10.3]	35.2 [-]
	Color	24.4 [23.8]	11.6 [11.7]	42.9 [42.8]
Accuracy congruency magnitudes	Total	4.5 [3.2]	6.2 [2.3]	27.9 [7.4]
	Grapheme	2.9 [1.7]	5.9 [2.7]	28.1 [10.4]
	Color	6.3 [4.8]	8.2 [5.4]	33.7 [19.8]
(C) Within-test change in IAT magnitudes				
RTs	Total	−3.8 [−2.5]	8.9 [7.3]	37.1 [25.3]
	Grapheme	−1.6 [0.5]	9.5 [8.4]	32.8 [31.6]
	Color	−5.8 [−4.4]	10.0 [8.5]	41.7 [26.8]
Accuracy	Total	2.8 [−0.8]	15.7 [3.3]	70.8 [12.9]
	Grapheme	5.2 [2.6]	12.0 [5.2]	55.7 [21.3]
	Color	0.2 [−4.6]	21.2 [5.4]	98.6 [18.3]