Cognition 186 (2019) 72–81

Contents lists available at ScienceDirect

Cognition

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

Original Articles

# Superior learning in synesthetes: Consistent grapheme-color associations facilitate statistical learning

Tess Allegra Forest[a,1], Alessandra Lichtenfeld[b,1], Bryan Alvarez[b], Amy S. Finn[a,*]

[a] Department of Psychology, University of Toronto, 100 St. George Street, 4th Floor, Sidney Smith Hall, Toronto, ON M5S 3G3, Canada
[b] Department of Psychology, University of California, Berkeley, Room 3210 Tolman Hall #1650, Berkeley, CA 94720-1650, USA

ARTICLE INFO

*Keywords:*
Learning
Synesthesia
Statistical learning
Language
Multimodal

ABSTRACT

In synesthesia activation in one sensory domain, such as smell or sound, triggers an involuntary and unusual secondary sensory or cognitive experience. In the present study, we ask whether the added sensory experience of synesthesia can aid statistical learning—the ability to track environmental regularities in order to segment continuous information. To investigate this, we measured statistical learning outcomes, using an aurally presented artificial language, in two groups of synesthetes alongside controls and simulated the multimodal experience of synesthesia in non-synesthetes. One group of synesthetes exclusively had grapheme-color (GC) synesthesia, in which the experience of color is automatically triggered by exposure to written or spoken graphemes. The other group had both grapheme-color and sound-color (SC+) synesthesia, in which the experience of color is also triggered by the waveform properties of a voice, such as pitch, timbre, and/or musical chords. Unlike GC-only synesthetes, the experience of color in the SC+ group is not perfectly consistent with the statistics that signal word boundaries. We showed that GC-only synesthetes outperformed both non-synesthetes and SC+ synesthetes, likely because the visual concurrents for GC-only synesthetes are highly consistent with the artificial language. We further observed that our simulations of GC synesthesia, but not SC+ synesthesia produced superior statistical learning, showing that synesthesia likely boosts learning outcomes by providing a consistent secondary cue. Findings are discussed with regard to how multimodal experience can improve learning, with the present data indicating that this boost is more likely to occur through explicit, as opposed to implicit, learning systems.

## 1. Introduction

Synesthesia is a perceptual phenomenon in which activation of one sensory or cognitive domain triggers a secondary experience that is involuntary, automatic, unusual, and consistent (Cytowic & Eagleman, 2009). Colored hearing is one of the most commonly studied varieties of synesthesia to date and is a broad label identifying synesthetes who experience colors triggered from natural and musical sounds, by letters and/or numerals, and/or by speech. Included in the latter category are grapheme-color synesthetes (GC synesthetes), who experience color induced from the *shape* of a grapheme (e.g., letter or number) whether it is printed on the page, spoken within the context of words or sentences[2] (Elias, Saucier, Hardie, & Sarty, 2003; Mattingley, Rich, Yelland, & Bradshaw, 2001), or imagined (Dixon, Smilek, Cudahy, & Merikle, 2000; Jansari, Spiller, & Redfern, 2006). Even though speech induces color in the absence of printed text, GC synesthetes' experience of color is linked to orthography, not phonology. Therefore whether spoken or written, the first component of the words cradle and cent would be the same color, while the initial component of sell and cent would not. A synesthete may also experience multiple forms of synesthesia, meaning that a synesthete may experience grapheme-color synesthesia and sound-color synesthesia, which is triggered by the properties of a sound, including things such as voice, pitch, timbre, musical chords and /or properties of speech.

Synesthetes have been the subject of several studies that attempt to understand if the added synesthetic perception can benefit learning and memory (Rothen, Meier, & Ward, 2012; Watson, Akins, Spiker, Crawford, & Enns, 2014). GC synesthetes have demonstrated better

* Corresponding author.
*E-mail address:* finn@psych.utoronto.ca (A.S. Finn).
[1] Authors contributed equally to this work.
[2] This is consistent with previous work in the area of psycholinguistics that graphemes are processed during speech comprehension, regardless of whether listeners are aware (36,37).

https://doi.org/10.1016/j.cognition.2019.02.003
Received 29 May 2017; Received in revised form 30 January 2019; Accepted 1 February 2019
Available online 11 February 2019

memory than controls for words that are presented both visually (Gross, Neargarder, Caldwell-Harris, & Cronin-Golomb, 2011; Radvansky, Gibson, & McNerney, 2011) and verbally (Mills, Innis, Westendorf, Owsianiecki, & McDonald, 2006; Rothen & Meier, 2009, 2010; Yaro & Ward, 2007). GC synesthetes also demonstrate better memory for colors (Yaro & Ward, 2007), simple abstract figures and shapes (Rothen & Meier, 2010), and concepts (Rothen & Meier, 2010). However, GC synesthetes show no memory advantage for numerical digits[3] (digit span; (Gross et al., 2011; Rothen & Meier, 2010)), digit matrices (Green & Goswami, 2008; Rothen & Meier, 2009; Yaro & Ward, 2007), or complex visual figures (Gross et al., 2011; Mills et al., 2006; Yaro & Ward, 2007) including faces (Gross et al., 2011). Overall, it appears that GC synesthetes perform better than controls when simple visual features, words, or speech-based information is to be remembered, but not necessarily for other types of information including individual graphemes (numbers/digits), and complex stimuli (such as faces).

These studies show that declarative—consciously recalled—memory can be enhanced in synesthetes. The same may also be true of learning construed more broadly (Watson et al., 2014). One study showed that novel associations (made with color) can transfer to existing color-based synesthetic associations (Meier & Rothen, 2007). Another study investigated artificial grammar learning using Markov chain grammars that were built over letters (synesthetically inducing) versus arbitrary symbols (synesthetically neutral). GC synesthetes significantly outperformed controls by identifying more grammatical strings, but only for strings comprised of letters; groups were equivalent when neutral symbols were used (Rothen et al., 2013). What's more, a further study discovered that synesthetes can learn rule-based categories using internally-generated synesthetic colors, outperforming non-synesthetes when no colors were physically present (Watson, Blair, Kozik, Akins, & Enns, 2012). Synesthesia can therefore benefit learning across a variety of circumstances.

It is currently unknown how synesthesia impacts statistical learning, a fundamental learning mechanism which allows learners to track the statistics present in their environment in order to segment continuous information into discrete, reliable units. Statistical learning is a widespread and powerful learning mechanism that was originally demonstrated in infants as a way to extract words from a stream of continuous speech when the reliable co-occurrence of syllables was the only cue present to help infants segment the speech (Saffran, Newport, & Aslin, 1996). The co-occurrence metric that is generally discussed in studies of statistical learning is transitional probability (or the likelihood of XY given an instance of X (Aslin, Saffran, & Newport, 1998)). However, statistical learning can operate over a number of other regularities and statistical relationships (Conway & Christiansen, 2005; Janacsek, Fiser, & Nemeth, 2012; Perruchet & Pacton, 2006; Schapiro, Turk-Browne, Norman, & Botvinick, 2015). Statistical learning can occur after very brief exposure (2 min or less (Batterink, 2017; Saffran, Aslin, & Newport, 1996)) and has been shown to operate across ages, species, and in multiple modalities (Campbell, Zimerman, Healey, Lee, & Hasher, 2012; Fiser & Aslin, 2001; Saffran et al., 1996; Toro & Trobalón, 2005). It has been put forward as a mechanism that allows learners to extract statistical structure from the environment in order to learn regularities that are foundational to language (Aslin & Newport, 2009), vision (Turk-Browne, Jungé, & Scholl, 2005), and action (Baldwin, Baird, Saylor, & Clark, 2003; Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015).

Of particular relevance for synesthesia, multimodal input has been shown to benefit statistical learning. This is true specifically when visual information (shape or color) is presented consistently at the level of the to-be segmented auditory unit. For example, a segmentation benefit is observed when each syllable of an auditory word is associated with the same visual shape (Cunillera, Laine et al., 2010; Thiessen, 2010) or color (Glicksohn & Cohen, 2013), when unique shapes are associated with unique sounds in a unit (Seitz, Kim, van Wassenhove, & Shams, 2007), and when visual information marks the boundary of auditory units (Cunillera, Camara et al., 2010a; Mitchel & Weiss, 2011). The multimodal experience of synesthetes is therefore likely to boost their statistical learning outcomes.

In the current study, we investigate exactly this. We ask whether the additional perceptual experience of GC synesthetes will facilitate statistical learning. In Experiment 1, we measure this directly in synesthetes. To gain insight into how the multimodal experience of synesthesia can benefit statistical learning, we further manipulate sound-color pairings during learning in non-synesthetes in Experiments 2 and 3.

## 2. Experiment 1

We predict that the additional perceptual experience of GC synesthetes will enhance statistical learning. An important question is whether synesthetes will show an advantage in segmenting both words and the categories to which those words belong. GC synesthetes will have both visual and auditory transitional probability cues to indicate (and segment) the words. For categories, however, there is no direct statistical cue in either the visual or auditory domains. It is therefore unknown whether the additional experience of color at the level of the word will enhance category learning. On the one hand, segmenting the words is an important first step to learning about categories and so learning those better could facilitate category learning. On the other hand, it has been shown that better learning of exemplars (in this case words) can make abstracting away from particular forms, and thus category learning, more difficult (Pitts Cochran, McDonald, & Parault, 1999).

Any advantage of GC synesthetes show should be due to the experience of color that is associated with hearing speech and having a consistent, redundant domain (vision, color) upon which to compute transitional probability. An advantage, however, could be due to some other factor associated with having synesthesia and the additional motivation to succeed on the part of participants who know they are being recruited because they have synesthesia. We therefore recruited another group of synesthetes that have both grapheme-color and sound-color synesthesia (the SC+ group), in which the experience of color is also triggered by the waveform properties of a voice, such as pitch, timbre, and/or musical chords and other properties of speech; while this group experiences color with speech (matching the domain of the experience of the GC group), the experience is not consistent with the statistics that signal word boundaries and is therefore not expected to benefit learning.

### 2.1. Materials and methods

#### 2.1.1. Participants

Thirty-one native-English speaking individuals with normal hearing and no reported history of neurological, psychiatric or learning disorders participated. Eleven made up the control group and were undergraduates from the University of California, Berkeley who participated for course credit (Mean age = 22.6 years, *SD* = 8.9, 8 female). Twenty were synesthetes who participated for monetary compensation, 10 were GC synesthetes (mean age = 21.3, *SD* = 8.3; 7 female) and 10 were SC+ synesthetes (mean age = 28.3, *SD* = 9.9; 9 female). GC synesthetes did not experience sound-color synesthesia (GC group), whereas the SC+ group experienced *both* sound-color and grapheme-color synesthesia.

Grapheme color synesthesia was confirmed for all synesthetes using the Synesthesia Battery (Eagleman, Kagan, Nelson, Sagaram, & Sarma,

[3] One explanation for this could be that numbers (for example 7) have both a grapheme and a word ("seven") that could be automatically associated with an auditorily presented digit. This could generate conflicting cues or possibly too many cues to leverage a learning advantage.

2007). Using previously established criteria (Carmichael, Down, Shillcock, Eagleman, & Simner, 2015), both groups were determined to be highly consistent in selecting the same color across the 3 iterative visual presentations of their triggering graphemes; the mean normalized distance across their choices in RGB (red, green, blue) color space was low for the GC group (0.78) and even lower for the SC+ group (0.68; Supplementary Fig. 1). For reference, any individual with a mean score below 1 is classified as a GC synesthete. (See Supplemental Materials for additional metrics from this battery.) Sound-color synesthesia was determined though personal interview in the lab. Participants were asked a series of questions to determine whether they had consistent sound-color associations that were elicited form any acoustic stimulus other than graphemes including voice, pitch, timbre, and musical chords.

#### 2.1.2. Procedure

Participants were asked to listen to the artificial language over headphones. Following previous work (Ettlinger, Finn, & Hudson Kam, 2011; Finn & Hudson Kam, 2008; Saffran, 2002; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), and to encourage passive listening, they were asked not to over-think or ignore what they were hearing and were permitted to draw anything other than graphemes to encourage this. After listening, participants completed two tests, the trials for which were presented randomly using Psychopy software (Peirce, 2007).

#### 2.1.3. Stimuli

Participants were exposed to an artificial language previously used to study word segmentation and grammatical category learning (Finn, Lee, Kraus, & Hudson Kam, 2014). The experimental stimuli lasted 9 min and 37 s and consisted of nine two-syllable words made from nine consonants and nine vowels. There were no silences or cues to word boundary other than transitional probability. Within words transitional probabilities were high (deterministic or 1.0) and low (0.33 or 0.5) across word boundaries. Each word fell into one of 3 categories (A, B and C; Fig. 1a). Category members had the same phonological structure and distribution: A words were created from a pattern of consonant (C)-vowel (V)-consonant (C)- vowel (V) sounds (CVCV), B words followed a CVVC pattern, and C words followed a CVCVC pattern. A words always followed C words, B words always followed A words and C always followed B.

As previously described (Finn et al., 2014), the artificial language stimuli and test items were generated with SoftVoice, a text-to-speech software that uses terminal analog formant synthesis (and not pre-recorded di-phones (Katz, 2005)). This was chosen to eliminate segmentation cues that were not experimentally relevant, holding co-articulation cues the same and keeping voice and pitch constant at a monotonic $F_0$ (fundamental frequency) of 83.62 Hz. Vowels in all of the stimuli and tests were the same length (170 ms) and consonants ranged from 60 ms to 140 ms (but were always the same for that phone regardless of their location).

#### 2.1.4. Tests

After exposure, participants were asked to complete word segmentation and grammatical category tests. All tests were two alternative forced choice in which participants were asked to listen to two items, separated by a 700 ms pause, and indicate which was more representative of the language they were just exposed to.

#### 2.1.5. Word segmentation

In the word segmentation test, participants were given a choice between a word and a foil. A word was one of the items listed in Fig. 1a (e.g. *mukeh*), which occurred in the stimuli and had a word-internal syllable-to-syllable transitional probability of 1.0. A foil consisted of either a non-word or a part-word. Non-words never occurred in the stimuli (thus having a syllable-to-syllable transitional probability of 0) and were made up of two syllables from separate words (e.g. *dukeh*). Part-words consisted of the second syllable of one word with the first syllable from a different word that had previously occurred after that word during exposure (e.g. *odti*). In the speech stream, these syllables occurred in succession, but had lower transitional probabilities than words (0.33 or 0.5 as opposed to 1.0). There were 18 items total, nine pitting words against non-words and nine pitting words against part-words.

#### 2.1.6. Category level test

For the category level test, learners chose between two three-word strings that can be thought of as sentences with no pauses. Test items made use of novel words in strings with familiar words from the exposure phase of the experiment. Eighteen novel words were generated, nine followed one of the three phonological category structures (three for each of the phonological categories; these are referred to as novel-good words; Supplementary Table 1), and nine did not follow the phonological structure of any of the A, B, or C words (referred to as novel-bad words; Supplementary Table 2).

Test items were also made of two main sub-types. In the first sub-type (*novel-good*) learners were asked to compare a full three-word "sentence" using a novel category congruent word (novel-good) that was in the correct place relative the other familiar words in the sentence to a different three-word sentence using a novel category congruent word (novel-good) that was in the *incorrect* place relative to the other words in the sentence. In the second sub-type (*novel-good versus bad*) learners were asked to compare a full three-word sentence using a novel category congruent word (novel-good) that was in the correct place relative the other familiar words in the sentence to a novel category incongruent word (novel-bad) in a sentence. (Note there is no correct placement for novel-bad words since they are not part of the categorical structure). Crucially, because novel words were created that had never appeared in exposure, there was no way that participants could use their knowledge of the transitional probabilities between syllables to choose the correct option. This means that any above-chance performance on these tests reflects participants having learned something about the underlying sound-based category structure. There were eight novel-good test items and six novel good versus bad. See Table 1 for examples.

#### 2.1.7. Data analysis

Analyses were conducted using R (R Core Team, 2012). Repeated measures analyses of variance (ANOVAs) were used to analyze omnibus effects pertaining to test-type. For tests with no repeating items (the category level test), one-sample *t*-tests were used to compare performance to chance. In tests with item repetition (the word segmentation test), generalized mixed-effects models, predicting accuracy, (glmer function in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015)) were used to assess whether performance differed from chance, whether learning occurred during test, and differences across learning groups accounting for item repetition during test. All models contained random intercepts and random slopes for the within-subject variables included as fixed effects, grouped by subject. Differences from chance were assessed in a model containing only an intercept as a fixed effect. Learning during test was assessed in a model with test item repetition and learning group (with the interaction term) included as fixed effects, and learning differences by group—accounting for repetition during test—were assed in a model containing test item repetition and learning group (with no interaction term). See the Supplementary Material for exact model specifications in R. All raw data are also available through the Supplementary Material (Forest, Lichtenfeld, Alvarez, & Finn, 2019).

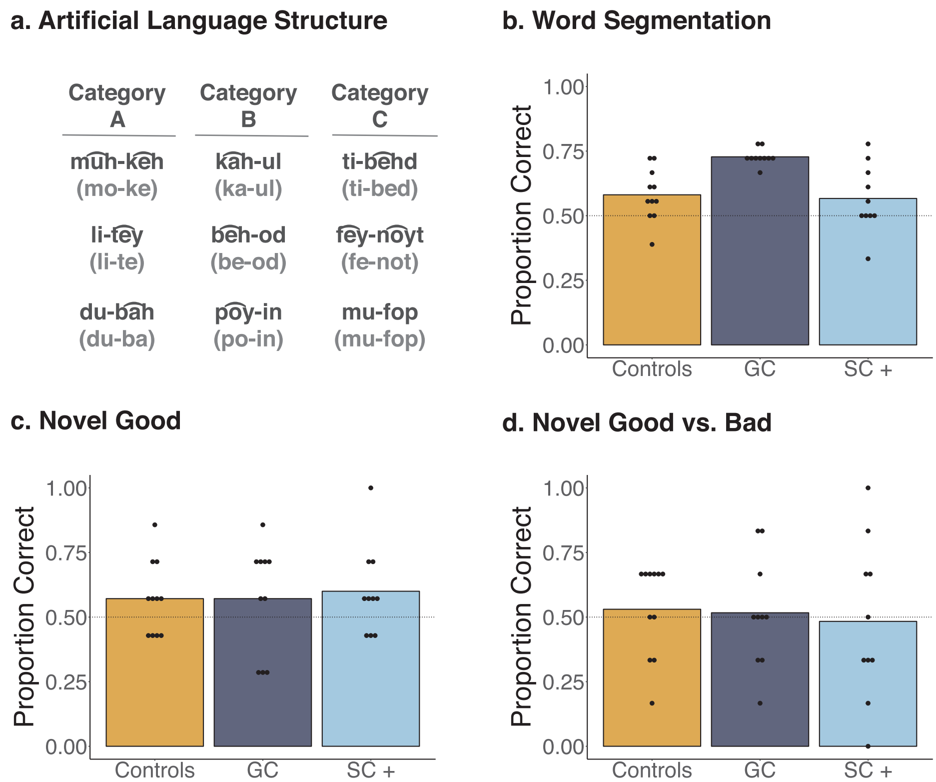


**Fig. 1.** *Experiment 1.* (a) Darker words are written phonetically, and connecting bars are meant to indicate how vowels were pronounced (i.e., the vowel '*o͡y*' is pronounced as the final sound in the English word "boy;" the vowel '*e͡h*' is pronounced as the middle sound in the English word "let.") Words in parentheses (gray) demonstrate the consonant-vowel structure of each word. The mean proportion correct is plotted for each learning group for: the word segmentation test (b), the novel-good test items (c), and novel-good-versus-bad test items (b). Controls are plotted in orange, GC synesthetes in dark blue, and SC+ synesthetes in light blue. Dots indicate the performance of individual participants and the dashed lines represent chance (50%).

## 2.2. Results

### 2.2.1. Word segmentation

A repeated measures ANOVA examining performance on the 3 types of word segmentation sub-tests (transitional probabilities of 0 vs. 1, 0.33 vs. 1, and 0.5 vs. 1) showed that performance did not differ across the sub-tests ($F(2, 56) = 1.81$, $p = .170$, $\eta_p^2 = 061$). An aggregate measure (average of all word segmentation sub-tests) was therefore used in all subsequent analyses. Across all groups learning was better than chance (Mean = 62%, $\beta = 0.25$, $z = 5.09$, $p = 3.52e-07$; Fig. 1b), a finding that was also true when only considering the first presentation of each word (Mean = 56%, $\beta = 0.31$, $z = 2.56$, $p = .01$).

Word repetition did not impact performance during test ($\beta = 0.37$, $z = 1.29$, $p = .196$), suggesting that learning did not occur during the test itself. In models that nonetheless accounted for word repetition during test, we observed that GC synesthetes performed better than both controls (Mean GC Synesthetes = 73%, Mean Controls = 58%, $\beta = 0.66$ $z = 2.99$, $p = .003$) and SC+ synesthetes (Mean SC+ Synesthetes = 57%, $\beta = 0.72$, $z = 3.19$, $p = 0.001$). Performance in the SC+ synesthetes and control groups did not differ ($\beta = -0.058$, $z = -0.28$, $p = .780$). Thus, while all groups demonstrated successful learning, GC synesthetes were superior.

### 2.2.2. Category level test

Across all learning groups, one-sample t-tests revealed that performance was significantly better than chance for the novel-good items, (Mean = 58%, $t(30) = 2.599$, $p = .014$, $d = .476$; Fig. 1c), but not the novel good versus bad items (Mean = 51%, $t(30) = 0.259$, $p = .798$, $d = 0.046$; Fig. 1d). Thus, subjects show knowledge of the category structure when one novel item is generated and tested, but not when two items, one that fits with a category form and one has a new form, are generated. A between-subjects ANOVA was conducted on the learned subtest (*novel-good*) and revealed no group differences ($F(2, 28) = 0.087$, $p = .917$, $\eta_p^2 = 0.006$; Fig. 1c). Thus, all groups showed learning of the categorical structure (as measured by the *novel good* items), but this did not differ based on whether one had synesthesia (or the type of synesthesia they have).

**Table 1**
Example test items.

| Definition | Example |
|---|---|
| *Word segmentation test* | |
| Word vs. *non-word* | Mukeh vs. *dukeh (transitional probability = 0)* |
| Word vs. *part word* | Beyod vs. *odti (transitional probability = 0.33)* |
| *Category level test* | |
| Novel-good correct vs. *incorrect* place | $\mathrm{A_{word}}$ $\mathbf{B_{novel\ (cvvc)}}$ $\mathrm{C_{word}}$ vs. $\mathrm{A_{word}}$ $\mathrm{C_{word}}$ $\mathbf{B_{novel\ (cvvc)}}$<br>mukeh **dehuwt** tibehd vs. mukeh tibehd **dehuwt** |
| Novel-good vs. *novel-bad* | $\mathrm{A_{word}}$ $\mathbf{B_{novel\ (cvvc)}}$ $\mathrm{C_{word}}$ vs. $\mathrm{A_{word}}$ $\mathbf{N_{novel\text{-}bad(cvccv)}}$ $\mathrm{C_{word}}$<br>mukeh **dehuwt** tibehd vs. mukeh **neytlae** tibehd |

### 2.3. Discussion

GC synesthetes outperformed both non-synesthetes and SC+ synesthetes on measures of word segmentation. Further, while it was possible that GC synesthesia could benefit the learning of grammatical categories (via better learning of the words), we did not observe any group differences on category learning. Below we discuss the possibilities as to why synesthetes do not differ from controls on the category learning measure and why GC synesthetes show superior word segmentation.

#### 2.3.1. Category level test

All groups demonstrated learning of the categorical structure as measured by performance on the *novel-good* test items and no group showed learning as measured by the novel *good versus bad test items*. Poor performance on novel *good versus bad items* is expected given previous work (Finn et al., 2014). As stated in that work, it is likely that learners are not sure what to do with the novel "bad" words, since these are comprised of a phonological structure that does not exist in the artificial language. All groups showed significant learning of the category structure based on performance on the *novel-good* test and performance did not differ across groups. Thus, GC synesthesia afforded no benefit to learning the categorical structure of the artificial language.

It was possible that *both* the GC and SC+ synesthetes could have learned the grammatical categories better than controls because synesthetes have a unique pattern of neural connectivity that could underlie an enhanced ability to draw connections among exemplars (Whitaker et al., 2014), something that is important for learning categories. It was also possible that because GC synesthetes showed superior word segmentation, that their grammatical category-level performance would also be better than the other groups. This, however, was not the case, fitting with previous work showing that better learning of words does not necessarily lead to better learning of the categories (Finn et al., 2014). In fact, there are demonstrations that learning the first level information well (in this case words) can actually prevent abstraction and thinking about how a given word fits within the broader structure (Gómez, 2002; Marchman, 1993; Pitts Cochran et al., 1999; Ramscar & Gitcho, 2007).

Importantly, synesthetes did not perform any differently from controls on this measure. Of note, this test does not measure participants' knowledge of the statistical information present in the stimuli, but rather the distributional patterns that govern the operation of words. The word segmentation test, on the other hand, *does* measure knowledge of statistical structure since it pits items that have high versus low transitional probability directly against one another.

#### 2.3.2. Word segmentation

We observe a very clear benefit in word segmentation for GC synesthetes. It is likely that the superior performance of GC synesthetes is *due* to their synesthetic concurrents providing an additional and consistent cue to word boundaries, thereby enhancing the statistical signal by making it present in both the visual and auditory modalities. Indeed, as reviewed above, we know that multimodal input can enhance statistical learning (Cunillera, Camara et al., 2010a; Glicksohn & Cohen, 2013; Seitz et al., 2007; Thiessen, 2010).

Importantly, however, the SC+ synesthetes also have synesthetic concurrents, but their learning is not better than controls. We think this is because the visual experience of SC+ synesthetes does not directly reflect the transitional probability structure of the auditory stimulus as it does in the GC only group, but we cannot measure this in syntheses directly. Thus, to gain more insight into the *cause* of the learning advantage in our GC group, we ran two additional experiments in which we directly manipulated the visual experience of non-synesthetes to match what we think is happening in GC and SC+ participants respectively.

## 3. Experiment 2

In Experiment 2 we therefore sought to emulate the experience of GC synesthetes in non-synesthetes by creating a perfect statistical match between a visually-displayed color and the auditory stimulus using the same language as in Experiment 1. We therefore paired each syllable with one and only one color so that the auditory and visual information matched exactly. After consultation with grapheme-color-only synesthetes, we chose to assign a unique color to each syllable. Since our stimuli are quite rapid (each syllable is 230–310 ms), GC-only synesthetes indicated experiencing a color at the level of the syllable that was unique – sometimes initiated by the first consonant or vowel and sometimes by the vowel.

Further, since we did not observe a difference across groups on the category level test in Experiment 1, we chose to leave this test out and enhance our understanding of word segmentation in various ways. First, to gain more insight into the nature of participants' learning—-whether they were aware of the learning they express—we modified the word segmentation test to also ask participants to rate the confidence of each of their choices. We also added an indirect measure of learning (Batterink, Reber, Neville, & Paller, 2015; Ryan, 2012) and measured how quickly learners were able to detect syllables that were predicted (occur in the second position of a word) as compared to syllables that were not predicted (occurring in the first position of a word). This combination of measures allows us to gain insight into participants' learned representations—that is, how implicit or explicit their knowledge is.

### 3.1. Materials and methods

#### 3.1.1. Participants

Sixty undergraduates at the University of Toronto participated in exchange for course credit. Thirty participated in the one-to-one-multimodal condition (mean age = 19.65 years, *SD* = 1.05; 21 female) and 30 participated in an auditory-only control group (mean age = 19.36 years, *SD* = 3.6; 22 female).

#### 3.1.2. Stimuli

Participants in the auditory-only group were exposed to the same experimental stimuli as in Experiment 1. Participants in the one-to-one-multimodal group were exposed to the same auditory information as participants in the audio-only group, but were also exposed to simultaneously presented visual information. This visual information was made up of large colored ovals ($29 \times 20$ cm) that appeared on a computer screen in perfect synchrony with the auditory presentation of syllables. Each of the 18 unique syllables was assigned a unique color (Fig. 2a; Supplementary Table 3). Thus, visual information had exactly the *same* statistical regularities as the auditory information.

#### 3.1.3. Tests

After exposure, participants were asked to complete word segmentation and target detection tests. The word segmentation test always took place before the target detection task.

*3.1.3.1. Word segmentation.* Participants in the auditory control group completed the same word segmentation test as in Experiment 1 with the one modification: they were asked to rate how confident they were about each decision on a four-point scale with '1' being not at all confident and '4' being very confident. Participants in the one-to-one-multimodal group completed the exact same test except that the color information present during exposure was also present during the test.

*3.1.3.2. Target detection.* For this task, participants were presented with one of the 18 syllables from the artificial language and asked to detect this syllable from a continuous stream of the artificial language as quickly as possible. Before detecting syllables, participants completed a

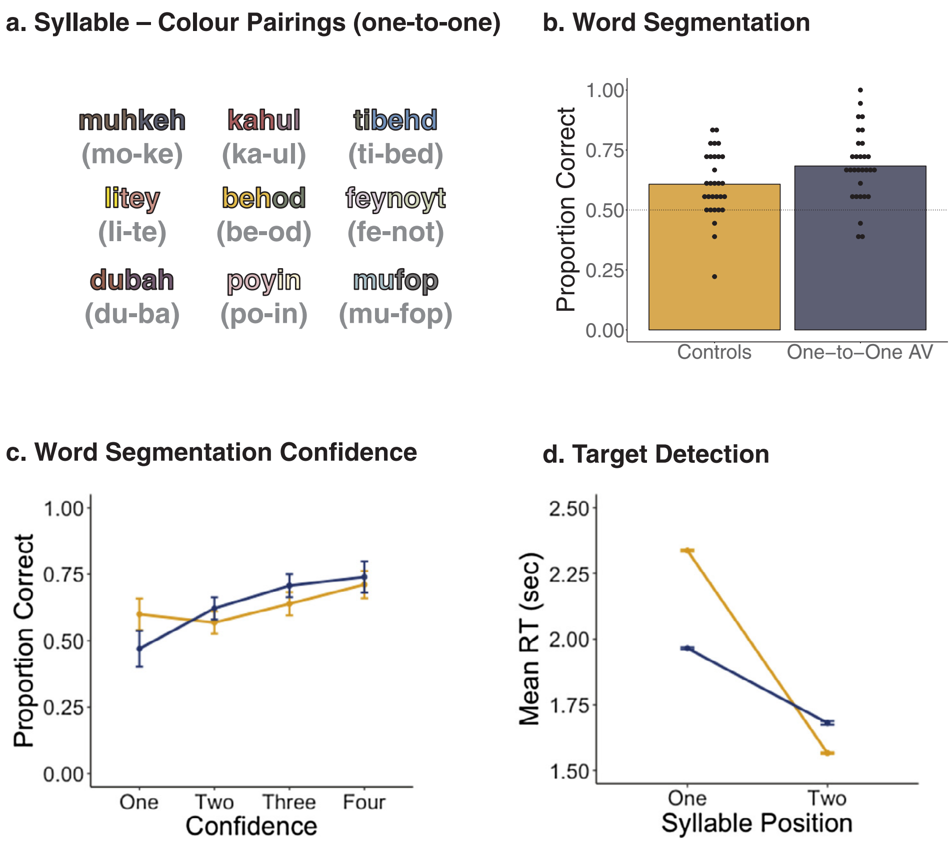


**Fig. 2.** *Experiment 2.* (a) Words presented in color are spelled phonetically with the color indicating the patch of color that was displayed in sync with that syllable (for the one-to-one-multimodal learners). Words in parentheses (gray) are written as they would be spelled in English. (b) The mean proportion correct is plotted for each learning group separately for the word segmentation test. Dots indicate the performance of individual participants and the dashed lines represent chance (50%). In this and all other graphs in this figure, controls are plotted in orange and one-to-one-multimodal learners are plotted in dark blue. (c) Mean proportion correct is plotted by confidence (one being not at all confident and four being very confident). Error bars represent standard error of the difference. (d) Mean reaction time is plotted by syllable position with one being the first (not predicted) syllable and two being the second (predicted) syllable. Error bars represent standard error of the difference.

practice stream in which they were provided with one of the syllables ('muh') which they would detect again during the test, and given the same instructions. Both instances of detecting this syllable were then excluded from further analyses. In all, participants were asked to detect all 18 syllables one at a time in mini-streams of the training stimulus that were approximately 14 s long and contained 2 repetitions of each word and three repetitions the target word. To promote detection, these short streams were slowed down to an average of 390 ms per syllable. A unique stream was generated for the detection of each syllable and the order of syllables was counterbalanced. The one-to-one-multimodal group completed this test with the addition of the visual information, while the auditory-only group completed the task with only the auditory information.

#### 3.1.4. Data analyses

For the word segmentation test, the same generalized mixed-effects models (glmer function in the lme4 package (Bates et al., 2015)) from Experiment 1 were used to assess whether performance differed from chance, whether learning occurred during test, and whether learning differed across learning groups accounting for item repetition during test.

Two additional models were run to (1) asses the relationship between confidence and accuracy on the word segmentation test, and (2) examine if reaction times during the target detection task differed for syllables that were predicted (second syllable) versus syllables that were not (first syllable). Both models included random intercepts and random slopes for the within-subject variables included as fixed effects, grouped by subject. A generalized mixed-effects model (glmer function in the lme4 package (Bates et al., 2015)) was used to assess the relationship between confidence and accuracy and included item repetition, confidence, and learning group as fixed effects. A Mixed-effects model (lmer function in lme4 package) was used to assess differences in syllable position by reaction time. Syllable position (first or second position) and learning group were included as fixed effects. Before analyses, reaction times (RTs) with filtered with the following specifications: RTs lower than 100 ms and that fell below or above three standard deviations from the by-participant, by-position (position-one or two syllable) mean were excluded. See Supplemental Material for full model specifications in R. All raw data are also available through the Supplemental Material (Forest et al., 2019).

### 3.2. Results

#### 3.2.1. Word segmentation

A repeated measures ANOVA examining performance on the 3 types of word segmentation sub-tests (transitional probabilities of 0 vs. 1.0, 0.33 vs. 1.0, and 0.5 vs. 1.0) showed that performance did not differ across the sub-tests ($F(2, 118) = 1.85$, $p = .162$, $\eta_p^2 = 0.03$). An aggregate measure (average of all word segmentation sub-tests) was therefore used in subsequent analyses. Across all groups, learning was better than chance (Mean = 65%, $\beta = 0.625$, $z = 7.25$, $p = 4.07e-13$; Fig. 2b), a finding that was also true when only considering the first presentation of each word (Mean = 63%, $\beta = 0.527$, $z = 5.40$, $p = 6.52e-08$). Moreover, confidence was positively associated with performance across participants even when accounting for the repetition of test items ($\beta = 0.312$, $z = 3.81$, $p = 0.0001$; Fig. 2c), an effect that did not differ by condition ($\beta = 0.158$, $z = 0.98$, $p = 0.325$). Performance on this particular test therefore reflects knowledge that participants are aware of.

We observed no difference in performance across the first and second presentations of words during test ($\beta = 0.011$, $z = 0.06$, $p = 0.953$), suggesting that word repetition did not impact performance on this test. In a model that nonetheless accounted for word repetition, we observed that one-to-one-multimodal participants performed better than auditory-only participants (Mean one-to-one-multimodal = 68%, Mean auditory-only = 61%, $\beta = 0.353$, $z = 2.11$, $p = .035$), an effect that did not interact with confidence ($\beta = 0.158$, $z = 0.99$, $p = 0.325$). Thus, both groups learned and were aware of their learning, but similar

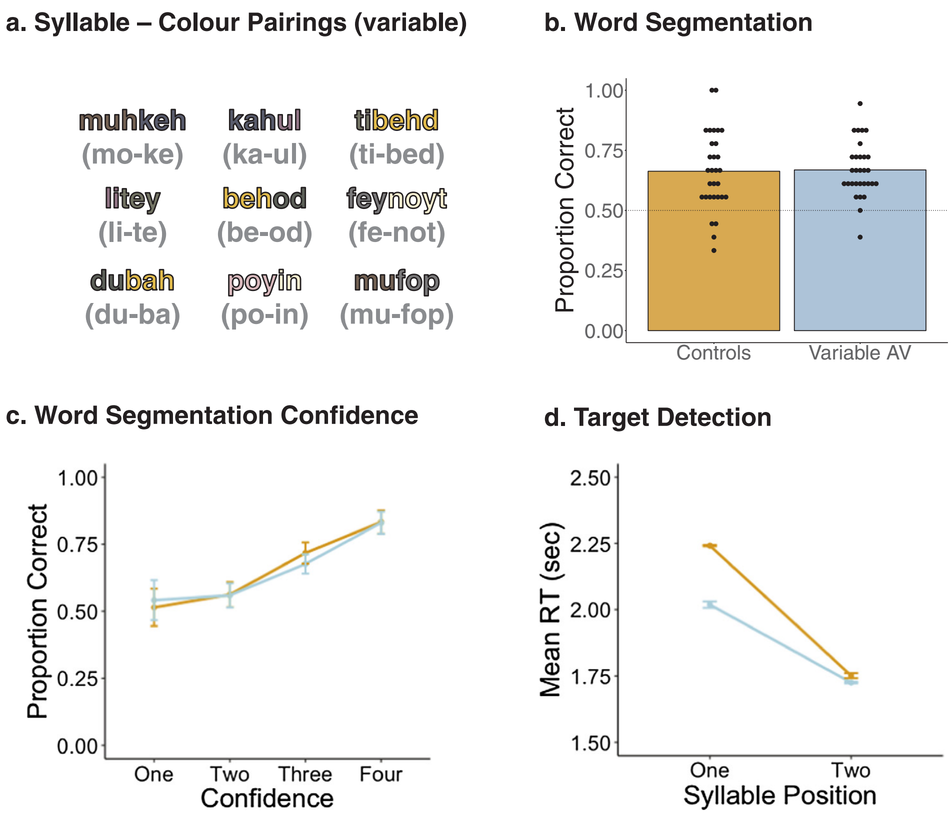


**Fig. 3.** *Experiment 3.* (a) Words presented in color are spelled phonetically with the color indicating the patch of color that was displayed in sync with that syllable (for the variable multimodal learners). Words in parentheses (gray) are written as they would be spelled in English. (b) The mean proportion correct is plotted for each learning group separately for the word segmentation test. Dots indicate the performance of individual participants and the dashed lines represent chance (50%). In this and all other graphs in this figure, controls are plotted in orange and the variable multimodal learners are plotted in light blue. (c) Mean proportion correct is plotted by confidence (one being not at all confident and four being very confident). Error bars represent standard error of the difference (d) Mean reaction time is plotted by syllable position with one being the first (not predicted) syllable and two being the second (predicted) syllable. Error bars represent standard error of the difference.

to the GC synesthetes in Experiment 1, the one-to-one-multimodal participants displayed a learning advantage.

#### 3.2.2. Target detection

Linear mixed-effects models show that participants, across learning groups, are faster to detect position two targets ($\beta = 0.514$, $t = 2.88$, $p = .004$; Fig. 2d), but that this does not interact with condition ($\beta = 0.487$, $t = 1.36$, $p = 0.174$). Thus, both groups show evidence of learning using this indirect measure, but the additional visual information does not impact performance on this test.

### 3.3. Discussion

Results from Experiment 2 show that one-to-one-multimodal learners outperformed auditory-only learners on the word segmentation measure, just as the GC synesthetes outperformed controls in Experiment 1. Of interest, the superior performance of these multimodal learners was restricted to the word segmentation task. Further, this experiment shows that participants are aware of the knowledge they demonstrate on the word segmentation test: confidence scales with performance. This suggests that the knowledge demonstrated on this test is explicit. Further, performance on the indirect, target detection measure, was indicative of learning, but did not differ by condition. Taken together, these data show that the superior performance of GC synesthetes in Experiment 1 is likely to be due to the consistent mapping between the auditory stimulus and visual concurrent, since a consistent mapping also produces superior learning in non-synesthetes. These data further show that this multimodal learning could benefit explicit representations (for which participants are aware) but not implicit representations (as measured indirectly via reaction times).

## 4. Experiment 3

While Experiment 2 showed that consistent audio-visual information could boost learning—even in non-synesthetes—it is unclear if the learning benefit has to do with the addition of *any* visual information or *consistent* visual information. Indeed, previous work has shown that multimodal exposure boosts learning, presumably by making to-be learned stimuli more salient (Shams & Seitz, 2008). In our final experiment, we therefore sought to understand how the consistency of the mapping impacts learning outcomes. In particular, we investigated whether a correlated but not perfect match between auditory and visual stimuli would improve learning. Previous work has only looked at perfectly correlated cross-modal information or randomly matched (completely unrelated) cross-modal information (Cunillera, Camara et al., 2010a; Glicksohn & Cohen, 2013; Seitz et al., 2007; Thiessen, 2010). We therefore sought to simulate the experience of the SC+ synesthetes in Experiment 1 who have multiple triggers for their synesthetic concurrents resulting in a more variable audio-visual mapping.

### 4.1. Materials and methods

#### 4.1.1. Participants

Sixty undergraduates at the University of Toronto participated in exchange for course credit. Thirty participated in the variable-multimodal condition (mean age = 19.88 years, *SD* = 3.1; 25 female) and 30 participated in an auditory-only control group (mean age = 19.37, *SD* = 1.3; years 20 female).

#### 4.1.2. Stimuli

Participants in the auditory-only group were exposed to the same experimental stimuli as in Experiments 1 and 2 (auditory-only group). Participants in the variable-multimodal group were exposed to the same auditory information as participants in the audio-only group, but were also exposed to simultaneously presented visual information. As in Experiment 2, this visual information was made up of large colored ovals (29 × 20 cm) that appeared on a computer screen in perfect synchrony with the auditory presentation of syllables. Instead of being assigned a unique color as in Experiment 2, the 18 unique syllables were assigned one of nine colors, (Fig. 3a; Supplementary Table 4).

Thus, the visual information has statistical structure that is correlated with, but not exactly matching the auditory structure. We built these variable mappings to match the experience of SC+ synesthesia in which there are various sources that could predict a similar percept.

#### 4.1.3. Tests

After exposure, participants were asked to complete word segmentation and target detection tests. These were the same as in Experiment 2; visual information was also present during test for the variable-multimodal group.

#### 4.1.4. Data analyses

All analyses and models were the same as in Experiment 2.

### 4.2. Results

#### 4.2.1. Word segmentation

A repeated measures ANOVA examining performance on the 3 types of word segmentation sub-tests (transitional probabilities of 0 vs. 1.0, 0.33 vs. 1.0, and.5 vs. 1.0) showed that performance differed across the sub-tests ($F(2, 116) = 4.19$, $p = .017$, $\eta_p^2 = 0.07$), with Tukey's posthoc pairwise tests indicating that performance did not differ on 0 vs. 1.0 and 0.33 vs. 1.0 items ($p = 0.47$) or 0 vs. 1.0 and 0.5 vs. 1.0 items ($p = .09$), but that performance on 0.33 vs. 1.0 items was higher than 0.55 vs. 1.0 ($p = .023$). Importantly, this main effect of test-type, did not interact with condition ($F(2, 116) = 1.65$, $p = .197$, $\eta_p^2 = 0.03$), so an aggregate measure (average of all word segmentation sub-tests) was used in subsequent analyses.

Across all groups, learning was better than chance (Mean = 67%, $\beta = 0.716$, $z = 8.47$, $p < 2e-16$); Fig. 3b), a finding that was also true when only considering the first presentation of each word (Mean = 61%, $\beta = 0.455$, $z = 4.55$, $p = 5.51e-6$). Moreover, confidence was positively associated with performance across participants even when accounting for the repetition of test items ($\beta = 0.631$, $z = 7.25$, $p = 4.03e-13$), an effect that did not interact with condition ($\beta = 0.068$, $z = 0.406$, $p = 0.685$). Performance on this particular test therefore reflects knowledge that participants are aware of.

We observed significantly better performance on the second presentations of words during test ($\beta = 0.571$, $z = 2.86$, $p = 0.005$), suggesting that word repetition during the test did impact performance in these participants. In a model that consequently accounted for word repetition, we observed that variable-multimodal participants performed no differently than auditory-only participants (Mean variable multimodal = 67%, Mean auditory-only = 66%, $\beta = 0.019$, $z = 0.115$, $p = 0.909$). Thus, both groups learned and were aware of their learning, but similar to the SC+ synesthetes in Experiment 1, the variable-multimodal participants did not differ from controls.

#### 4.2.2. Target detection

Linear mixed-effects models show that participants, across learning groups, are faster to detect position two targets ($\beta = 0.387$, $t = 2.25$, $p = 0.025$; Fig. 3c), but that this does not interact with condition ($\beta = 0.196$, $t = 0.57$, $p = 0.57$). Thus, both groups show evidence of learning using this indirect measure, but the additional visual information does not impact performance on this test.

### 4.3. Discussion

Results from Experiment 3 show that variable multimodal learners are no different from auditory-only learners on both the word segmentation measure and the target detection measure. Both groups display substantial learning on both measures, but this learning does not differ by condition, suggesting that the variable, but correlated color mapping is not sufficient to produce the learning boost observed in both Experiments 1 and 2.

## 5. General discussion

In the present study, we showed that synesthetes could outperform controls on a classic measure of statistical learning—an auditory word segmentation task. Specifically, we showed that individuals with grapheme-color-only (GC-only) synesthesia outperformed both non-synesthetes and synesthetes who have grapheme-color *and* sound-color synesthesia (the SC+ group). This is likely because the visual concurrents for GC-only synesthetes come from only one source and the concurrents for the SC+ synesthetes come from multiple sources, some of which are linked to the statistical relationship of interest and some of which are not.

To understand if this was the mechanism driving our observed group differences, we ran two additional experiments in non-synesthetes: one in which the visual and auditory information was matched perfectly (the one-to-one-multimodal group; Experiment 2) and one in which the mapping between the visual and auditory information was variable (the variable-multimodal group; Experiment 3). We found that learning was improved only when auditory and visual information matched perfectly. Thus, the learning advantage in GC-only synesthetes appears to be due to their consistent cross-modal mapping. Below, we discuss the implications of these findings for improving our understanding of both synesthesia and statistical learning.

### 5.1. Synesthesia

To our knowledge this is the first study to show that statistical learning outcomes can differ in synesthetes. This discovery is important. Statistical learning is thought of as a foundational learning system that supports our understanding of how the world is structured, whether these be regularities in our visual, motor, or auditory worlds (Aslin & Newport, 2012; Smith, Jayaraman, Clerkin, & Yu, 2018). The fact that this foundational learning system can operate differently in synesthetes is central to understanding how synesthetes extract regularities from their worlds.

Interestingly, it has been suggested that *all* infants have a form of synesthesia; the neonatal-synesthesia hypothesis suggests that all infants who experience stimulation in one sense can experience additional stimulation in another sense (Maurer, 1993; Wagner & Dobkins, 2011). While much remains to be discovered about this possibility, this suggests that *all* infant statistical learners might have different perceptual experiences (as compared with adults and children) which could impact their learning, quite possibly resulting in better learning under some circumstances. Much work is needed to know if these experiences would be in line with the one-to-one-multimodal and GC synesthetes here, but this possibility merits consideration in current thinking about the ways statistical learning changes with development (Gómez, 2017; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Smith et al., 2018).

### 5.2. Statistical learning

We join previous work in showing that statistical learning can benefit from multimodal input. Most of this previous work, however, assigns one visual stimulus to *all* elements of an auditory unit, using, for example, the same shape to mark each syllable of a "word" (Cunillera, Camara et al., 2010a; Cunillera, Laine et al., 2010; Thiessen, 2010). This is quite different from what we did by manipulating color at same level of the syllable. Two previous experiments did do something similar to this. One manipulated color at the level of the syllable (Glicksohn & Cohen, 2013), but had 24 unique syllables and only eight unique colors. They did not find a learning advantage in this condition, likely because of the variable mapping, as we also found in Experiment 3. The other experiment assigned unique shapes to unique artificial sounds (not syllables (Seitz et al., 2007)) and found a multimodal learning advantage. Taken together, this pattern of data is similar to ours. When the mappings are consistent (one-to-one) and not variable, learning outcomes are enhanced.

In our data, we show no learning advantage in SC+ synesthetes nor in our variable-mapping learning group, suggesting that noise in multimodal mappings is neither helpful nor harmful to unimodal learning. This would not have to be the case. It could be that one consistent color-auditory syllable (for example our syllable “poy” which is the only one to occur in pink in Experiment 3) would boost the signal helping to learn the whole word “poyin” and provide an “anchor” so to speak in helping to segment the rest of the words, as has been observed in unimodal studies of statistical learning in which a known word is included in the artificial language (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Cunillera, Camara et al., 2010b; Johnson & Tyler, 2010). We did not observe this, suggesting that there is something unique about multimodal input. Further studies are needed to probe the precise conditions under which multimodal input can benefit statistical learning.

Importantly, the variable mappings and experience of SC+ synesthetes also did not harm learning (as compared with the control and auditory-only groups). This suggests that there is no real cost to having variable or less consistent multimodal information, although the boundaries of this could be explored much further and work investigating how multiple levels of information are learned in statistical learning studies is greatly needed (c.f., (Brady & Oliva, 2008; Emberson & Rubinstein, 2016; Turk-Browne, Isola, Scholl, & Treat, 2008). It could be that when multiple modalities indicate a statistical relationship, they sum to produce a stronger learning signal and boost learning. When the signals are not as strongly correlated, however, they could be treated as separate therefore not adding to produce a stronger signal. Future experiments are needed to explore these ideas.

The present work is also important for thinking about how information that is learned via statistical learning is represented in the mind. Statistical learning is generally discussed as an automatic and *implicit* process that occurs with little effort or awareness on the part of the learner (Aslin & Newport, 2009; Saffran et al., 1997; Shohamy & Turk-Browne, 2013; Turk-Browne, Scholl, Chun, & Johnson, 2009). However, various studies have shown that attention can be important and even sometimes necessary for statistical learning (Finn et al., 2014; Musz, Weber, & Thompson-Schill, 2015; Toro, Sinnett, & Soto-Faraco, 2005; Vickery, Park, Gupta, & Berryhill, 2017). Indeed, it is increasingly clear that statistical learning is not process pure, tapping into both explicit and implicit learning systems (Batterink et al., 2015; Bays, Turk-Browne, & Seitz, 2016). We echo this here: participants across all conditions in Experiments 2 and 3 were able to display learning of the statistical structure using separate tests that tap into both explicit and implicit representations. Performance on the word segmentation test appears to be explicit since it scales with confidence. And since the target-detection-task maps subtle differences in reaction time based on whether an item is predicted, performance on this measure is likely to reflect implicit representations. Importantly, the benefit of one-to-one-multimodal learners was *only* observed on the word segmentation test, suggesting that the additional and consistent visual information present in this condition boosts learning through more explicit learning systems. This is important for furthering our understanding of *how* synesthesia and multimodal input can boost learning, suggesting that declarative or explicit learning systems and their associated neural circuitry are impacted most.

This work has therefore answered an important question at the crossroads of perception and learning: does multisensory experience boost statistical learning in synesthesia? Using a variety of tools to measure the nature of learned representations, we found that *yes* multisensory experience can boost statistical learning in synesthetes and non-synesthetes alike. Mechanistically, we further determined that the learning boost is likely due to consistent cross-modal pairing and that the observed learning benefit is more likely to occur though *explicit* learning systems.

## Acknowledgements

This work was funded in part by the National Institutes of Health (NRSA to A.S.F). We thank our insightful participants, Carla L. Hudson Kam and Katherine Duncan for their thoughts and comments on drafts of this work and Zahra Abolghasem and Gloria Rebello for help with data collection.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.02.003.

## References

Aslin, R. N., & Newport, E. L. (2009). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.). *Infant pathways to language: Methods, models, and research disorders* (pp. 15–29). New York, NY US: Psychology Press.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*, 170–176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2003). Infants parse dynamic action. *Child Development, 72*, 708–717.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 1*.

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological Science, 28*, 921–928.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory & Language, 83*, 62–78.

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2016). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition, 23*, 1072–1097.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298–304.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science, 19*, 678–685.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology, 76*, 30–77.

Campbell, K. L., Zimerman, S., Healey, M. K., Lee, M. M. S., & Hasher, L. (2012). Age differences in visual statistical learning. *Psychology and Aging, 27*, 650–656.

Carmichael, D. A., Down, M. P., Shillcock, R. C., Eagleman, D. M., & Simner, J. (2015). Validating a standardised test battery for synesthesia: Does the Synesthesia Battery reliably detect synesthesia? *Consciousness and Cognition, 33*, 375–385.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology/Learning, Memory & Cognition, 31*, 24–39.

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010b). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57*, 134–141.

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010a). Speech segmentation is facilitated by visual cues. *Quartely Journal of Experimental Psychology, 63*, 260–274.

Cunillera, T., Laine, M., Càmara, E., & Rodríguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language, 63*, 295–305.

Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia.* Cambridge, MA: MIT Press.

Dixon, M. J., Smilek, D., Cudahy, C., & Merikle, P. M. (2000). Five plus two equals yellow. *Nature, 406* 365–365.

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*, 139–145.

Elias, L. J., Saucier, D. M., Hardie, C., & Sarty, G. E. (2003). Dissociating semantic and perceptual components of synaesthesia: Behavioural and functional neuroanatomical investigations. *Cognitive Brain Research, 16*, 232–237.

Emberson, L. L., & Rubinstein, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition, 153*, 63–78.

Ettlinger, M., Finn, A. S., & Hudson Kam, C. L. (2011). The effect of sonority on word segmentation: Evidence for the use of a phonological universal. *Cognitive Science, 36*, 655–673.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*, 477–499.

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When It hurts (and helps) to try: The role of effort in language learning. *PLoS One, 9*, e101806.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499–504.

Forest, T. A., Lichtenfeld, A., Alvarez, B., & Finn, A. S. (2019). Synesthesia and statistical learning dataset. *Cognition Supplementary Material* [dataset].

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review, 20*, 1161–1169.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.
Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 1–9.
Green, J. A. K., & Goswami, U. (2008). Synesthesia and number cognition in children. *Cognition, 106*, 463–473.
Gross, V. C., Neargarder, S., Caldwell-Harris, C. L., & Cronin-Golomb, A. (2011). Superior encoding enhances recall in color-graphemic synesthesia. *Perception, 40*, 196–208.
Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science, 15*, 496–505.
Jansari, A. S., Spiller, M. J., & Redfern, S. (2006). Number synaesthesia: When hearing "four plus five" looks like gold. *Cortex, 42*, 253–258.
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339–345.
Katz, J. (2005). *SoftVoice (demo program)*. Los Angeles, CA: SoftVoice Inc.
Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*, 215–234.
Mattingley, J. B., Rich, A. N., Yelland, G., & Bradshaw, J. L. (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. *Nature, 410*, 580–582.
Maurer, D. (1993). Neonatal synesthesia: Implications for the processing of speech and faces. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.). *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 109–124). Dordrecht: Springer, Netherlands.
Meier, B., & Rothen, N. (2007). When conditioned responses "fire back": Bidirectional cross-activation creates learning opportunities in synesthesia. *Neuroscience, 147*, 569–572.
Mills, C. B., Innis, J., Westendorf, T., Owsianiecki, L., & McDonald, A. (2006). Effect of a synesthete's photisms on name recall. *Cortex, 42*, 155–163.
Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1081–1091.
Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception, & Psychophysics, 77*, 78–96.
Peirce, J. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8–13.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Science, 10*, 233–238.
Pitts Cochran, B., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language, 41*, 30–58.
Radvansky, G. A., Gibson, B. S., & McNerney, M. W. (2011). Synesthesia and memory: Color congruency, von Restorff, and false memory effects. *Journal of Experimental Psychology Learning Memory & Cognition, 37*, 219–229.
Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences, 11*, 274–279.
R Core Team (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing URL: http://www.R-project.org/.
Rothen, N., & Meier, B. (2009). Do synesthetes have a general advantage in visual search and episodic memory? A case for group studies. *PLoS ONE, 4*, e5037.
Rothen, N., & Meier, B. (2010). Grapheme–colour synaesthesia yields an ordinary rather than extraordinary memory advantage: Evidence from a group study. *Memory, 18*, 258–264.
Rothen, N., Meier, B., & Ward, J. (2012). Enhanced memory ability: Insights from synaesthesia. *Neuroscience & Biobehavioral Reviews, 36*, 1952–1963.
Rothen, N., Scott, R. B., Mealor, A. D., Coolbear, D. J., Burckhardt, V., & Ward, J. (2013). Synesthetic experiences enhance unconscious learning. *Cognitive Neuroscience, 4*, 231–238.
Ryan, J. D. (2012). Reconsidering the use of "explicit" and "implicit" as terms to describe task requirements. *Cognitive Neuroscience, 3*, 211–212.
Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.
Saffran, J. R., Aslin, R. N., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.
Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 1–15.
Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2015). Statistical learning of temporal community structure in the hippocampus. *Hippocampus, 26*, 3–8.
Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36*, 1445–1453.
Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General, 142*, 1159–1170.
Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences, 22*, 325–336.
Thiessen, E. D. (2010). Effects of Visual information on adults' and infants' auditory statistical learning. *Cognitive Science, 34*, 1093–1106.
Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*, B25–34.
Toro, J. M., & Trobalón, S. J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics, 67*, 867–875.
Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 399–407.
Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology, 134*, 552–564.
Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934–1945.
Vickery, T. J., Park, S. H., Gupta, J., & Berryhill, M. E. (2017). Tasks determine what is learned in visual statistical learning. *Psychonomic Bulletin & Review*.
Wagner, K., & Dobkins, K. R. (2011). Synaesthetic associations decrease during infancy. *Psychological Science, 22*, 1067–1072.
Watson, M. R., Akins, K., Spiker, C., Crawford, L., & Enns, J. (2014). Synesthesia and learning: A critical review and novel theory. *Frontiers in Human Neuroscience*, 8.
Watson, M. R., Blair, M. R., Kozik, P., Akins, K. A., & Enns, J. T. (2012). Grapheme-color synaesthesia benefits rule-based category learning. *Consciousness and Cognition, 21*, 1533–1540.
Whitaker, K. J., Kang, X., Herron, T. J., Woods, D. L., Robertson, L. C., & Alvarez, B. D. (2014). White matter microstructure throughout the brain correlates with visual imagery in grapheme–color synesthesia. *NeuroImage, 90*, 52–59.
Yaro, C., & Ward, J. (2007). Searching for Shereshevskii: What is superior about the memory of synaesthetes? *The Quarterly Journal of Experimental Psychology, 60*, 681–695.