A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Research Report

Colored-Speech Synaesthesia Is Triggered by Multisensory, Not Unisensory, Perception

Gary Bargary,[1,2,3] Kylie J. Barnett,[1,2,3] Kevin J. Mitchell,[2,3] and Fiona N. Newell[1,2]

[1]School of Psychology, [2]Institute of Neuroscience, and [3]Smurfit Institute of Genetics, Trinity College Dublin

ABSTRACT—*Although it is estimated that as many as 4% of people experience some form of enhanced cross talk between (or within) the senses, known as synaesthesia, very little is understood about the level of information processing required to induce a synaesthetic experience. In work presented here, we used a well-known multisensory illusion called the McGurk effect to show that synaesthesia is driven by late, perceptual processing, rather than early, unisensory processing. Specifically, we tested 9 linguistic-color synaesthetes and found that the colors induced by spoken words are related to what is perceived (i.e., the illusory combination of audio and visual inputs) and not to the auditory component alone. Our findings indicate that color-speech synaesthesia is triggered only when a significant amount of information processing has occurred and that early sensory activation is not directly linked to the synaesthetic experience.*

Synaesthesia occurs when a stimulus triggers an anomalous percept in the same or different modality as the normal percept. Reported synaesthetic experiences have ranged from tasting words (Ward & Simner, 2003) and sounds (Beeli, Esslen, & Jäncke, 2005) to seeing calendar units (Smilek, Callejas, Dixon, & Merikle, 2006), but by far the most studied form of synaesthesia is induction of color by letters, words, or digits (Rich & Mattingley, 2002). This latter type of synaesthesia is diversely referred to as colored speech (Baron-Cohen, Harrison, Goldstein, & Wyke, 1993), colored hearing (Marks, 1975), or linguistic-color (Simner, Glover, & Mowat, 2006), lexical-color (Rich, Bradshaw, & Mattingley, 2005), or grapheme-color (Ramachandran & Hubbard, 2001) synaesthesia; such inconsistent terminology reflects an underlying lack of understanding about the amount of information processing required for synaesthesia to be induced. For example, several studies have found that synaesthesia can occur very rapidly (Palmeri, Blake, Marois, & Whetsell, 2002; Ramachandran & Hubbard, 2001; Smilek, Dixon, Cudahy, & Merikle, 2001) and is sensitive to changes in low-level properties of the inducing stimulus, such as contrast (Hubbard, Manoha, & Ramachandran, 2006) or font (Witthoft & Winawer, 2006). These findings suggest that synaesthesia is an automatic association driven by early, unisensory input. However, attention, semantic information, and feature-binding processes (Dixon, Smilek, Duffy, Zanna, & Merikle, 2006; Esterman, Verstynen, Ivry, & Robertson, 2006; Mattingley, Payne, & Rich, 2006; Mattingley, Rich, Yelland, & Bradshaw, 2001; Muggleton, Tsakanikos, Walsh, & Ward, 2007; Myles, Dixon, Smilek, & Merikle, 2003; Rich & Mattingley, 2003; Sagiv, Heer, & Robertson, 2006) have also been implicated in mediating synaesthesia; such findings suggest that higher-level stages of information processing are required before the synaesthetic experience is elicited.

To assess whether synaesthesia is triggered by early, unisensory input or by relatively later perceptual processing, we used incongruent audiovisual recordings of spoken words known to induce the McGurk effect (McGurk & MacDonald, 1976). This effect occurs when incongruent visual information (i.e., a viseme) influences the perceived auditory information (i.e., spoken word), resulting in a change in the perceived sound from the real auditory input. For example, if the audio word /bait/[1] is dubbed onto the viseme of a person saying "gate" (i.e., [gate]), then, in the majority of cases, an observer will perceive the word *date*, a percept that is the result of the integration of the two sensory inputs.

Although the McGurk effect is considered to be the result of integration between the visual and auditory inputs, it is

Address correspondence to Fiona N. Newell, Institute of Neuroscience, Lloyd Building, Trinity College, Dublin 2, Ireland, e-mail: fiona.newell@tcd.ie.

[1]Slashes (e.g., /cat/) indicate auditory words, and brackets (e.g., [cat]) indicate visemes.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

attributed to processing in higher-level, heteromodal cortical areas rather than lower-level sensory areas (Calvert, Campbell, & Brammer, 2000; Hasson, Skipper, Nusbaum, & Small, 2007). In particular, activation in lower-level cortical areas involved in the processing of stimulus properties, such as the primary auditory cortex, was reported to be either unrelated to the perception of the illusion (Hasson et al., 2007) or subsequently related to feedback from higher-level, frontal cortical areas involved in speech production (Skipper, van Wassenhove, Nusbaum, & Small, 2007). Thus, the McGurk effect is thought to involve relatively late perceptual processing, rather than early sensory processing.

We predicted that if synaesthesia is triggered by early unisensory input, then the color induced by an incongruent audiovisual speech event should be related to the color induced by the auditory component presented in isolation. However, if the synaesthetic color is triggered at a later stage of information processing (e.g., after integration of the heard and seen speech), then the induced color should be related to what is perceived (i.e., the illusory or "fused" percept), rather than the color induced by the auditory component alone.

METHOD

Participants

Twelve synaesthetes (1 male, 11 female; mean age = 43.9 years) took part in the experiment for pay. They experienced colors when they perceived graphemes and were previously tested for consistency in their grapheme-to-color associations (Barnett et al., 2007). All participants reported normal or corrected-to-normal vision and no hearing abnormalities.

Stimuli and Apparatus

Our stimuli were based on audiovisual recordings of a male speaker (full-face view), which were made using a digital video camera (JVC, Model GR-DVL167). The actor was instructed to speak individual words at a normal pace, and each recorded sample was 2 s long. These words were selected on the basis of viseme and phonetic properties known to give rise to McGurk illusions: The place of oral articulation of the visual word should be more posterior in the mouth than that of the dubbed auditory word in order for the illusion to occur (Alsius, Navarra, Campbell, & Soto-Faraco, 2005). Some combinations of words were expected to result in the perception of an entirely new word (e.g., the auditory word /bait/ combined with the viseme [gate] would result in the illusory perception of the word *date*); in combinations in which the place of articulation of the visual and auditory word were close together, the perceived word was expected to match the viseme (e.g., the auditory word /bent/ combined with the viseme [dent] would result in the perception of the word *dent*).

Each audiovisual sample was edited using Adobe Premiere 6.0 software for the PC. We first created a basic set of 32 stimuli for the incongruent, audiovisual (McGurk) condition by splicing together the audio version of one word with the viseme of another word. Audio and visual samples were paired in a nonrandom manner, constrained by the place of articulation of the critical phoneme, such that the phoneme had a more posterior place of articulation for the viseme (e.g., /gate/, /tail/, /nap/) than for the auditory word (e.g., /bait/, /pale/, /map/, respectively; for a similar procedure, see Alsius et al., 2005). Pilot studies using a large set of stimuli in which the spoken word was not congruent with the viseme allowed us to select a set of words that had a high likelihood of inducing a McGurk illusion (for a list of the experimental word set used in this study, see Table S1 in the supporting information available on-line; see p. 533).

The visual-only and audio-only stimuli were subsequently created from this incongruent, audiovisual set of stimuli. Visual-only stimuli were created by masking the incongruent audio component (i.e., the actor's spoken word) with white noise. For the audio-only stimuli, we adopted a previously reported technique (Alsius et al., 2005; Campbell & Massaro, 1997) in which we masked the viseme using spatial quantization with a mask measuring 10 pixels horizontally and 15 pixels vertically over the face image (see Fig. 1).

Participants were seated 57 cm away from the main computer monitor, and each visual stimulus subtended a visual angle of 18°. Auditory stimuli were presented at an intensity of 60 dB using two loudspeakers fixed to either side of the monitor. All auditory, visual, and audiovisual stimuli were presented using Presentation software (Neurobehavioral Systems, Inc., Albany, CA).

On each auditory, visual, or audiovisual trial, participants were simultaneously presented with a color array on a different monitor, which was positioned adjacent to the main monitor. Custom-built software allowed this color array to be interactive such that the participant could select a specific, individual color from a wide array using a computer mouse. Chosen colors were automatically recorded and specified in RGB space.

Design

The experiment was based on a one-way, within-subjects design with three different conditions: auditory, visual, and audiovisual (incongruent). Trials in each condition were presented in separate blocks, and block order was counterbalanced across participants. Each block contained 32 trials: The first two were practice trials (and were excluded from the data analysis), and the remaining 30 were presented in random order across participants.

Procedure

A 1,000-ms fixation cross marked the beginning of each trial. After fixation, and according to block type, an auditory, visual, or incongruent, audiovisual stimulus appeared for 2,000 ms. The participant's task was to verbally report what word had been spoken by the actor as accurately as possible (and if in doubt, to

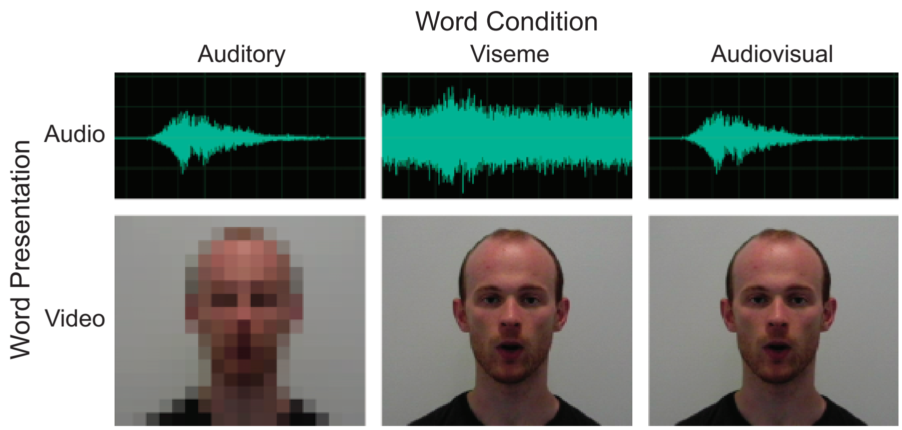

Fig. 1. Sample stimuli showing a male actor articulating words and the relevant audio channel. To create the incongruent, audiovisual (i.e., McGurk) stimuli, the audio channel was artificially dubbed over an incongruent video channel. For the auditory condition, a spatial quantization technique was used to mask the visual information. Likewise, in the visual-only (viseme) condition, white noise was used to mask the auditory signal.

give his or her best guess), and the experimenter recorded the participant's response. Before reporting the word, however, the participant was instructed to choose the color on the array that best matched the color induced by the word he or she had heard.

For audiovisual trials, we used previously reported criteria to classify perceptions as either illusory or not (Alsius et al., 2005). Specifically, we determined that the illusion did not occur if the auditory component of the word was reported. A McGurk illusion had occurred if the reported word was different from either the auditory or the visual component of the word, or if the reported word matched the viseme, provided the viseme was not recognized on its own. Reported words that did not fall into these categories were labeled as "other."

RESULTS

Although all our participants reported experiencing colors when viewing graphemes before our study, 3 reported that no color was induced by any of the experimental word stimuli presented aurally. The remaining 9 synaesthetes all reported synaesthetic colors in each condition (see Fig. 2 for an example).

On average, participants reported experiencing the McGurk illusion in 21.8 (or 73%) of the total 30 incongruent, audiovisual trials. In the remaining incongruent, audiovisual trials, either the auditory (5.3 trials) or an unrelated (2.9 trials) word was reported. In the auditory condition, participants reported the correct word in 19.6 of the 30 trials, on average. For the remaining trials, a McGurk (7.2 trials) or unrelated (3.2 trials) word was reported. Finally, few visemes were correctly identified (i.e., on average, less than 1 out of 30 trials); in other cases, an unrelated word (16.8 trials), a related word in which the first letter was the same (4.2 trials), or no word (8.1 trials) was reported. Interestingly, synaesthetic colors were reported only when a subsequent word was also reported, which suggests that viseme information alone was not sufficient to induce a synaesthetic color in the absence of perceiving a word.

The McGurk illusion was experienced and the word was reported correctly in the auditory condition for an average (across 9 participants) of 13.2 out of 30 words ($SD = 3.2$). However, we excluded cases in which the same color was induced by the illusory word and auditory-only word if these words contained different letters in their critical phonemes (e.g., if *m* and *n* induced blue) or if these words shared the same first letter (because synaesthetes subsequently reported that the first letter of a word typically induced the synaesthetic color). This left an average of 6.8 words ($SD = 3.2$), for which we computed the average RGB vector distance between the colors induced by the McGurk illusion and those induced by the correctly reported auditory-only counterpart (for further details, see Ward, Huckstep, & Tsakanikos, 2006). The mean vector distance across participants was 212.3 ($SD = 80.6$; or 151.09 on all 13.2 trials).

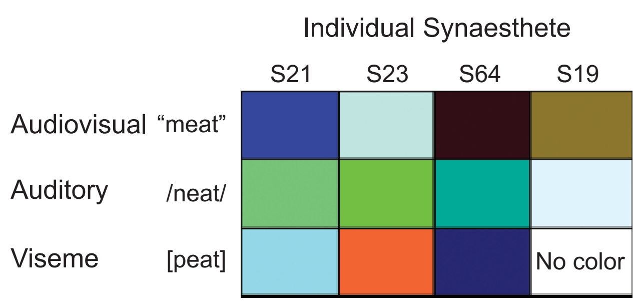

Fig. 2. Example of the different synaesthetic colors reported by 4 synaesthetes (S21, S23, S64, and S19) in the auditory, viseme, and audiovisual conditions. The auditory word /neat/ and the viseme [peat] were correctly reported by these synaesthetes. The colors induced by these stimuli are shown (with the exception of S19, who did not report any color for this stimulus in the viseme condition). The figure also shows the colors induced when the auditory word /neat/ was combined with the viseme [peat]. In this case, the synaesthetes perceived the illusory audiovisual word *meat*, and the colors induced by the illusory word differed from those induced by the unisensory words alone.

We then compared this score against a baseline RGB vector distance, which was measured as the distance between the colors induced by the incongruent, audiovisual conditions when the illusion was not experienced and the colors induced by the correctly reported auditory-only counterpart, that is, the vector distance between the induced colors when the same word was perceived in the two conditions (3.9 out of 30 trials). This baseline RGB-distance score was 41 (SD = 16.51), averaged across participants. Participants sometimes reported the same illusory word in both the incongruent, audiovisual condition and the auditory condition (5.1 of the 30 trials); in these cases, the average RGB vector-distance score was 49 (SD = 18.02), which was close to our baseline measure. Because these two baseline RGB vector-distance scores (41 and 49) were not statistically different, $t(16) = 1.10$, $p = .33$, we pooled them together and obtained an average baseline measure of 47.5, which represented the variation in reported colors when the same word was perceived in both the audiovisual condition and the auditory condition (a sum of 9 trials). We then compared this baseline measure to the average RGB vector color distance of 212.3, calculated from the trials when the illusion was experienced, and found that they were significantly different, $t(16) = 6.04$, $p < .001$; this comparison was also significant for the vector distance of 151.09, $t(16) = 5.08$, $p < .001$. Furthermore, this significant difference (Bonferroni corrected for multiple comparisons) was consistently found for each of the individual synaesthetes except one, for whom a nonsignificant trend was observed (see Fig. 3): Unlike the other synaesthetes, participant S17 had a very limited synaesthetic color spectrum (she reported experiencing only three different colors in response to all inducing stimuli), resulting in many instances where similar colors were reported for different words.

To verify our baseline measure, we conducted an experiment with 5 of our synaesthetes, to whom we presented only congruent audiovisual stimuli in which the perceived words matched the illusory words perceived in the incongruent, audiovisual conditions of the earlier experiment. For example, if *date* was previously perceived from the incongruent combination of /bait/ and the viseme [gate], then we presented *date* both visually and aurally in a congruent audiovisual trial. The average RGB vector distance between the colors induced by these congruent-audiovisual trials and the colors induced by the incongruent, audiovisual trials was 60.2. This distance was not significantly different from our previously calculated baseline measure of 50.8 observed for the same 5 synaesthetes, $t(8) = 1.14$, $p = .28$, which suggests that there was no difference between the colors induced by the same words, regardless of whether they were illusory or not.

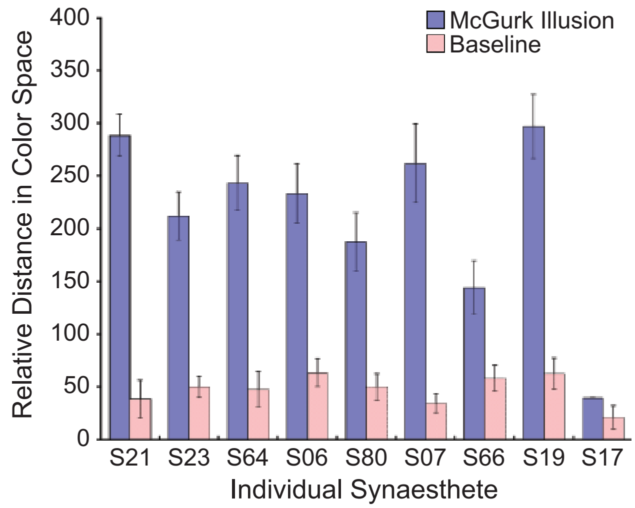

Fig. 3. Plot showing the mean RGB vector distances between the colors reported in the audiovisual condition and in the auditory condition across participants. The baseline difference was calculated as the difference between the audiovisual condition and the auditory condition when the same word was reported in both (i.e., when no McGurk illusion was experienced). The McGurk difference was calculated as the difference between the audiovisual condition and the auditory condition when different words were reported (i.e., when the McGurk illusion was experienced). Error bars indicate standard errors of the means.

DISCUSSION

The results indicated that the colors induced by the McGurk illusion (i.e., colors triggered by incongruent, audiovisual conditions) were different from the colors induced by the auditory-only components of the stimuli. Thus, the colors induced by audiovisual words are related to what is perceived, and not to early sensory processing. Furthermore, multisensory integration occurred before triggering the synaesthetic association; that is, visual information from the speaker's face had already become integrated with the acoustic information before the synaesthetic color was induced. This result indicates that synaesthetic colors to speech are induced by late perceptual processing, rather than early unisensory processing.

The finding that the synaesthetic color is driven by relatively high-level perception rather than early unisensory information may provide support for theories of synaesthesia involving feedback from higher-order areas to early visual areas, such as color areas (Dixon et al., 2000; Ward et al., 2006). Recent evidence that audiovisual speech is represented at an abstract level, independent of its sensory components (Hasson et al., 2007), further supports this model. However, it also remains possible that audiovisual (or indeed auditory) inducers may be reencoded into a visual form (grapheme), and this reencoded information triggers a synaesthetic concurrent via direct cross-activation (Hubbard, 2007). This reencoding may not be a common mechanism, or may be insufficient in some cases to trigger the color experience, because 3 of our participants in the main study did not report experiencing colors to any of our aurally presented word stimuli. Our data clearly suggest that, in the case of synaesthesia induced by spoken words, lower-level sensory properties of the stimuli are not tied to the induced color, but that a more abstract, multisensory representation can elicit the synaesthetic experience.

Acknowledgments—This research was funded by an Irish Research Council for Science, Engineering, and Technology scholarship (to G.B.) and by Health Research Board Grant RP/2004/19 (to F.N.N. and K.J.M.). We thank Phil Merikle (University of Waterloo, Waterloo, Ontario, Canada) for providing the color-array software.

REFERENCES

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*, 839–843.

Barnett, K.J., Finucane, C., Asher, J.E., Bargary, G., Corvin, A.P., Newell, F.N., & Mitchell, K.J. (2007). Familial patterns and the origins of individual differences in synaesthesia. *Cognition*, *106*, 871–893.

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

Beeli, G., Esslen, M., & Jäncke, L. (2005). Synaesthesia: When coloured sounds taste sweet. *Nature*, *434*, 38.

Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649–657.

Campbell, C., & Massaro, D. (1997). Perception of visible speech: Influence of spatial quantization. *Perception*, *26*, 129–146.

Dixon, M.J., Smilek, D., Duffy, P.L., Zanna, M.P., & Merikle, P.M. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex*, *42*, 243–252.

Esterman, M., Verstynen, T., Ivry, R.B., & Robertson, L.C. (2006). Coming unbound: Disrupting automatic integration of synesthetic color and graphemes by transcranial magnetic stimulation of the right parietal lobe. *Journal of Cognitive Neuroscience*, *18*, 1570–1576.

Hasson, U., Skipper, J.L., Nusbaum, H.C., & Small, S.L. (2007). Abstract coding of audiovisual speech: Beyond sensory representation. *Neuron*, *56*, 1116–1126.

Hubbard, E. (2007). Neurophysiology of synesthesia. *Current Psychiatry Reports*, *9*, 193–199.

Hubbard, E.M., Manohar, S., & Ramachandran, V.S. (2006). Contrast affects the strength of synesthetic colors. *Cortex*, *42*, 184–189.

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

Mattingley, J.B., Payne, J.M., & Rich, A.N. (2006). Attentional load attenuates synaesthetic priming effects in grapheme-colour synaesthesia. *Cortex*, *42*, 213–221.

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.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *265*, 746–748.

Muggleton, N., Tsakanikos, E., Walsh, V., & Ward, J. (2007). Disruption of synaesthesia following TMS of the right posterior parietal cortex. *Neuropsychologia*, *45*, 1582–1585.

Myles, K.M., Dixon, M.J., Smilek, D., & Merikle, P.M. (2003). Seeing double: The role of meaning in alphanumeric-colour synaesthesia. *Brain and Cognition*, *53*, 342–345.

Palmeri, J., Blake, R., Marois, F., & Whetsell, W. (2002). The perceptual reality of synesthetic colors. *Proceedings of the National Academy of Sciences, USA*, *99*, 4127–4131.

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

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

Rich, A.N., & Mattingley, J.B. (2002). Anomalous perception in synesthesia: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, *3*, 43–52.

Rich, A.N., & Mattingley, J.B. (2003). The effects of stimulus competition and voluntary attention on colour-graphemic synaesthesia. *NeuroReport*, *14*, 1793–1798.

Sagiv, N., Heer, J., & Robertson, L. (2006). Does binding of synesthetic color to the evoking grapheme require attention? *Cortex*, *42*, 232–242.

Simner, J., Glover, L., & Mowat, A. (2006). Linguistic determinants of word colouring in grapheme-colour synaesthesia. *Cortex*, *42*, 281–289.

Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., & Small, S.L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*, 2387–2399.

Smilek, D., Callejas, A., Dixon, M.J., & Merikle, P.M. (2006). Ovals of time: Time-space associations in synaesthesia. *Consciousness and Cognition*, *16*, 507–519.

Smilek, D., Dixon, M.J., Cudahy, C., & Merikle, P.M. (2001). Synaesthetic photisms influence visual perception. *Journal of Cognitive Neuroscience*, *13*, 930–936.

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.

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

Witthoft, N., & Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex*, *42*, 175–183.

(Received 8/8/08; Revision accepted 11/8/08)

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.