Functional and structural basis of the color-flavor incongruency effect in visual search

Jianping Huang¹, Fei Wang¹, Jie Sui², & Xiaoang Wan¹

1. Department of Psychology, Tsinghua University, Beijing, China
2. Department of Psychology, University of Bath, Bath, UK

Correspondence to Xiaoang Wan, Mingzhai 219, Tsinghua University, Beijing, China 100084. Telephone: 86-10-62796746. Fax: 86-10-62773687. Electronic mail: wanxa@mail.tsinghua.edu.cn
ABSTRACT

We conducted a functional magnetic resonance imaging (fMRI) study and a voxel-based morphometry (VBM) study to investigate the functional and structural basis of how visual search for flavor labels in packaging is influenced by the color of the packaging. The participants were cued by a flavor word before searching for a package with this flavor label. The behavioral results of both studies revealed that the participants were slower to find the target when its color was incongruent with the flavor label in terms of color-flavor associations than when it was congruent with the flavor label, which is indicative of a color-flavor incongruency effect in the reaction times (RTs). The fMRI results revealed that this behavioral effect was associated with enhanced activation in the right putamen. The VBM results further revealed a significant positive correlation between the magnitude of the behavioral effect and volume of gray matter in the right putamen. Taken together, these findings suggest that the color-flavor incongruency effect may be attributed to the violation of color expectation in the incongruent condition and that the putamen may be one of the important areas for processing events in violation of expectation.

Keywords: expectation; color-flavor associations; incongruency; packaging
1. Introduction

People tend to associate an experience of stimuli, objects, or events in one sensory modality with the experience in another sensory modality, a phenomenon referred to as crossmodal correspondences (Spence, 2011; Spence & Deroy, 2013). Crossmodal correspondences can be developed between a pair of stimuli or features based on the commonalities in their neural coding, their repeated co-appearance in everyday life, common language to describe them, and/or common affective responses they may elicit (Spence, 2011; Wan et al., 2014a). In particular, the crossmodal correspondences between color and flavor, namely, color-flavor associations (for a review, see Spence et al., 2015), have been observed with food (e.g., Levitan, Zampini, Li, & Spence, 2008; Zhou et al., 2015), beverages (Wan et al., 2014b, 2016), and pharmaceutical pills (Wan et al., 2015). Based on the repeated exposure to the packaging of mainstream products with large market shares, consumers might even associate a certain packaging color with a particular flavor for packaged food such as potato chips (for a review, see Spence & Velasco, 2018) and, therefore, develop explicit and/or implicit associations between packaging colors and flavor labels (Piqueras-Fiszman & Spence, 2011; Piqueras-Fiszman, Velasco, & Spence, 2012).

Such packaging color-flavor label associations might be set up based on the color-flavor association or the semantic congruence (i.e., both mapped to a specific identity or meaning) between the packaging color and flavor label via associative learning (see Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008; Knoeferle,
Knoeferle, Velasco, & Spence, 2016; Marks, 1978; Parise & Spence, 2013). Specifically, when participants were cued by a flavor word to search for such a label in packaging in the subsequent display, they were slower to find the target when its color was incongruent with the participants’ packaging color-flavor label associations (i.e., a congruent target) than when it was a congruent target (Velasco et al., 2015). Here, we refer to this effect as the color-flavor incongruency effect.

Why does this color-flavor incongruency effect occur in visual search for food labels? Velasco et al. (2015) attributed it to volitional control based on the participants’ expectations concerning the color of the target that guides attention (see Theeuwes, 2010). A flavor word such as “tomato-flavor” might activate the participants’ relevant representations of this flavor and generate expectations concerning the color of the target packaging based on their packaging color-flavor label associations. If this is the case, the task of searching for a flavor label could be temporarily transformed to searching for an expected color, presumably because searching for a color singleton is much simpler than searching for a target word (e.g., Dampuré, Ros, Rouet, & Vibert, 2012). Therefore, an incongruent target would violate the participants’ expectation and could only be found after the participants have switched back to the less-efficient word search. We refer to this possibility as the violation of expectation account.

Alternatively, the color-flavor incongruency effect might also be explained by attention captured by the color being primed by the semantic representations of the
object mentioned in the flavor word. Importantly, a word expressing color could automatically prime attention capture by stimuli displayed in this color (Ansorge & Becker, 2012). Object-color associations can be so strong (Palmer & Schloss, 2010) that a word expressing an object can prime attention being captured by pictures or words displayed in the color associated with this object (Huettig & Altmann, 2011; Léger & Chauvet, 2015). Therefore, it is possible that the semantic representations of the object mentioned in the flavor word might prime attention capture by the color associated with this object; for example, the word “tomato” in “tomato-flavor” might prime attention capture by the color red. Given the consistency between natural object-color pairings and packaging color-flavor label associations (Velasco et al., 2014), a color-flavor congruent target has the color being primed for attention capture, resulting in an advantage in reaction times (RTs) compared to that of the incongruent target. We refer to this possibility as the primed attention-capture account.

The violation of expectation and primed attention-capture accounts both predict the RT difference between the congruent and incongruent target conditions, so they could hardly be differentiated by behavioral data without substantial modifications to the task or the experimental paradigm. Therefore, we conducted a combined fMRI and VBM study to differentiate these two accounts by investigating the neural basis of the color-flavor incongruency effect. Specifically, we aimed to use the functional imaging data to differentiate the violation of expectation and primed attention-capture accounts at the neural level and then to explore whether variations of brain structures
associated with the color-flavor incongruency effect provide converging evidence with the functional imaging data.

2. An fMRI study of the color-flavor incongruency effect

In the present study, we used an orthogonal design of Target Congruency × Distractor Congruency, so that the violation of expectation and primed attention-capture accounts would make different predictions on the imaging results. Such an experimental design resulted in four experimental conditions, including the congruent target-congruent distractor, incongruent target-congruent distractor, congruent target-incongruent distractor, and incongruent target-incongruent distractor conditions (see Table 1 for a summary).

<table>
<thead>
<tr>
<th>Target</th>
<th>Distractor</th>
<th>Color expectation</th>
<th>Primed color</th>
<th>Attention capture by color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>Congruent</td>
<td>Fulfilled</td>
<td>Displayed in target</td>
<td>Yes</td>
</tr>
<tr>
<td>Incongruent</td>
<td>Congruent</td>
<td>Violated</td>
<td>Absent</td>
<td>No</td>
</tr>
<tr>
<td>Congruent</td>
<td>Incongruent</td>
<td>Fulfilled</td>
<td>Displayed in target</td>
<td>Yes</td>
</tr>
<tr>
<td>Incongruent</td>
<td>Incongruent</td>
<td>Violated</td>
<td>Displayed in distractor</td>
<td>Yes</td>
</tr>
</tbody>
</table>

One of the critical comparisons we planned to conduct was between the congruent target-incongruent distractor and incongruent target-incongruent distractor conditions. While the color being primed for attention capture is always present in both conditions, participants’ expectations concerning the target color would be fulfilled in the congruent target-incongruent distractor condition because the target
would be congruent but would be violated in the incongruent target-incongruent distractor condition because the target would be incongruent (see Table 1 for summary and Figure 1 for illustrations). Therefore, the violation of expectation account would predict significant differences in the activations between these two conditions in areas linked to the violation of expectation (i.e., striatal structures such as the putamen; for meta-analyses, see D’Astolfo & Rief, 2017; Garrison, Erdeniz, & Done, 2013). By contrast, the primed attention-capture account would predict no significant differences in the functional imaging data between these conditions.

<table>
<thead>
<tr>
<th>Congruent target-</th>
<th>Incongruent target-</th>
<th>Congruent target-</th>
<th>Incongruent target-</th>
</tr>
</thead>
<tbody>
<tr>
<td>congruent distractor</td>
<td>congruent distractor</td>
<td>incongruent distractor</td>
<td>incongruent distractor</td>
</tr>
</tbody>
</table>

Figure 1. Sample illustrations of four experimental conditions in searching for a target with the tomato flavor label. Specifically, panel (A) shows a congruent target among congruent distractors, an incongruent target among congruent distractors, a congruent target among incongruent distractors, and an incongruent target among incongruent distractors.
incongruent distractors, from the left to the right. In panel (B), we highlighted the target on each display with a circle, and the expected color (red) was absent in the incongruent target-congruent distractor conditions but present in the other three conditions.

The other critical comparison we planned to conduct was between the incongruent target-congruent distractor and incongruent target-incongruent distractor conditions. As can be seen in Figure 1, the color being expected and primed for attention capture would be absent in the incongruent target-congruent distractor condition, resulting in a violation of color expectation without attention capture by this color. By contrast, the color being expected and primed for attention capture would be presented in one of the distractors in the incongruent target-incongruent distractor condition, so the participants’ expectations would be violated, but their attention would be captured by the primed color. In other words, participants’ color expectation would be violated in both conditions, while their attention would be captured by the primed color in the incongruent target-incongruent distractor condition but not in the incongruent target-congruent distractor condition. Therefore, the violation of expectation account would predict no significant differences in the functional imaging data between these two conditions; in contrast, the primed attention-capture account would predict significant differences in the activation of brain areas related to attention capture, such as the bilateral intraparietal sulcus (IPS), bilateral frontal eye fields, and the right ventral-frontal cortex (for comprehensive reviews, see Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002).
2.1 Method

Participants

Forty Chinese participants (mean age = 20.7 ± 2.31 years, ranging from 18 to 27 years; 20 male and 20 female) took part in this study. All of the participants in this fMRI study and the following VBM study reported to be right-handed, and to have normal or corrected-to-normal visual acuity without color blindness or any neurological history. Both studies were approved by the Human Research Ethics Committee of Center for Biomedical Imaging Research of Tsinghua University, and performed in accordance with the ethical standards laid out in the Declaration of Helsinki. All participants were recruited from the participant pool of the Psychology Department of Tsinghua University. They gave written informed consent before the experiment started, and received 100 Chinese Yuan for their time and participation.

Apparatus & stimuli

The present study was conducted at the Center of Biomedical Imaging Research, Tsinghua University. A Pentium-based computer installed with the E-prime 2.0 software (Psychology Software Tools, Pittsburg, PA, USA) was used to run the study in a testing room outside the scanner room. Inside the scanner room, the stimuli were presented via a LCD projector onto a rear projection screen at the head end of the
scanner, and viewed via an angled mirror positioned attached to the head-coil at a
distance of approximately 60 cm.

Similar to Velasco et al.’s (2015) first experiment, each visual search display in
the present experiment consisted of three images of potato chip packaging (each
subtending 8.57° horizontally and 13.96° vertically) with a fictional brand “Crispies”
against a white background. The flavor label (subtending 4.49° horizontally and 1.46°
vertically) for each packaging consisted of three Chinese characters. On each search
display, three of the four quadrants of the screen were randomly chosen to present
these three images, and each image was centered on one quadrant. Three different
images were quasi-randomly chosen for each trial based on whether their color and
flavor combinations were congruent with the color-flavor associations reported by
Velasco et al. (2014). Specifically, the congruent stimuli were tomato flavor in red (H:
0°, S: 68%, B: 83%), cucumber flavor in green (H: 102°, S: 51%, B: 69%), chicken
flavor in orange (H: 21°, S: 67%, B: 91%), and BBQ flavor in burgundy (H: 350°, S:
75%, B: 59%). The incongruent stimuli were cucumber flavor in red, tomato flavor in
orange, chicken flavor in burgundy, and BBQ flavor in green. Prior to each search
display, a flavor word consisting of three Chinese characters (subtending 6.67°
horizontally and 1.9° vertically) was presented at the center of a white screen (see
Figure 2A for illustrations).
Figure 2. (A) An illustration of a trial in the fMRI study. (B) A flowchart of the block design for the fMRI study. Note that Rest = rest block, CtCd = congruent target-congruent distractor condition, ItCd = incongruent target-congruent distractor condition, CtId = congruent target-incongruent distractor condition, and ItId = incongruent target-incongruent distractor condition.

**Design and procedure**

As described in the Introduction section, we used a 2 (Target Congruency: congruent or incongruent) × 2 (Distractor Congruency: congruent or incongruent)
within-participants design, resulting in four conditions, including the congruent target-congruent distractor, incongruent target-congruent distractor, congruent target-incongruent distractor, and incongruent target-incongruent distractor conditions.

We used a block-design to present a total of 192 trials, evenly divided across the four experimental conditions. Specifically, there were 32 experimental blocks in total, and each block contained 6 trials of the same condition. As shown in Figure 2B, the scan was divided into 4 runs, and each run consisted of 8 experimental blocks, with 1 resting block (lasting 24 s) being inserted after every 4 experimental blocks. The order of the blocks within each run was quasi-randomized, so each type of trials was presented once in the first 4 blocks and once in the last 4 block. The total scanning time was approximately 16 minutes. Before the scanning started, each participant completed 12 practice trials in a resting room outside.

As can be seen in Figure 2A, each trial started with a 1000-ms white screen with a centered fixation cross (1.3° × 1.3°) in black, followed by a 1000-ms white screen with a flavor word presented at the center. Then another fixation screen was presented for 500 ms; followed by a search display for a maximum of 1500 ms. The participants were instructed to use different fingers to press different keys on the 4-key response box to indicate the location of the target as quickly and accurately as possible. Specifically, they were instructed to use their left middle, left index, right middle, or right index finger to press the top-left, bottom-left, top-right, or bottom-right key on
the response box when the target was presented in the corresponding quadrant of the screen, respectively.

*Image acquisition*

Images were acquired with a Philips Achieva 3.0T scanner with a standard 8-channel head coil. Functional data were acquired in an interlaced fashion along the AC-PC line with a T2-weighted EPI sequence of 37 transverse slices (TR = 2300 ms, TE = 35 ms, flip angle = 90°, acquisition matrix = 96 mm × 94 mm) of 5 mm in thickness, with 1 mm inter-slice gap. Within each session, a total of 110 EPI images were acquired. A High-resolution T1-weighted image was also acquired for each participant with 160 contiguous sagittal slices of 1 mm thickness and 8° flip angle. The sensitivity encoding (SENSE) parallel-imaging technique was employed to reduce the scan time (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999). The SENSE acceleration factor was 2 for the anterior-posterior direction and 1.5 for the right-left direction. Repetition time was set to be 8.2 ms, whereas echo time was 3.8 ms. The acquisition matrix was 256 voxels × 256 voxels × 160 voxels with a voxel size of 0.938 mm × 0.938 mm × 1 mm.

*Imaging data analysis*

We used SPM8 (Wellcome Department of Cognitive Neurology, London, UK) in MATLAB R2013b (Mathworks, Natick, MA, USA) to analyze the imaging data. First, all functional images were slice-timing, realigned, and co-registered with the
structural images. And the six estimated movement parameters were added as covariates in the 1st level model. Then we normalized these images to the Montreal Neurological Institute (MNI) template with a voxel size of 3×3×3 mm³ and smoothed them with a Gaussian kernel of 8 mm full-width at half-maximum to decrease spatial noise.

As we described earlier, we defined bilateral putamen, IPS, frontal eye fields, and the right ventral frontal cortex as the ROIs for the present study. We obtained the coordinates of putamen from the anatomical automatic labeling atlas (AAL; Tzourio-Mazoyer et al., 2002), and used the coordinates of the IPS, frontal eye fields, and the right ventral frontal cortex employed by Fox et al. (2006). We performed the general linear model (GLM) to analyze the preprocessed imaging data, which generated a statistical parametric map for each participant for each of the four conditions, convolved with the canonical hemodynamic response function (HRF). Then we used MarsBar (http://marsbar.sourceforge.net/) to extract beta values from a 10-mm radius sphere around the coordinates of ROIs, and performed group-level voxel-based pairwise comparisons between the congruent target-incongruent distractor and incongruent target-incongruent distractor conditions, and between the incongruent target-congruent distractor and incongruent target-incongruent distractor conditions.

We also performed Target Congruency ×Distractor Congruency repeated-measure Analyses of Variance (ANOVAs) within the whole brain, followed
by planned pairwise comparisons between the congruent target-incongruent distractor and incongruent target-incongruent distractor conditions, and between the incongruent target-congruent distractor and incongruent target-incongruent distractor conditions. Statistical maps were thresholded at $p < 0.01$. We addressed the issues of multiple comparisons with a cluster extent threshold of 719 voxels (equivalent to corrected $p < 0.05$), which was estimated by the 3dClustSim program of AFNI software (Cox, 1996) via the auto-correlation function (ACF) approach (ACF values: 0.59, 6.39, 18.93; 5000 iterations).

2.2 Results

*Behavioral results*

In the present study, the participants showed a high level of accuracy of 91.37% overall. We excluded RTs which were three standard deviations shorter or longer than the group means from the following data analyses, resulting in 4.92% of the data being discarded. The mean RTs based on correct trials and accuracy data for each condition are shown in Figure 3.
Figure 3. Mean RTs (in ms) and accuracy in the four experimental conditions of the fMRI study. Note that * denotes $p < 0.05$, *** denotes $p < 0.001$, and the error bars show the standard errors of the means.

The 2 (Target Congruency: congruent or incongruent) × 2 (Distractor Congruency: congruent or incongruent) repeated-measure ANOVAs revealed a significant main effect of Target Congruency on the RTs, $F(1, 39) = 217.18$, $p < 0.001$, $\eta^2_p = 0.85$, and on the accuracy data, $F(1, 39) = 8.94$, $p = 0.005$, $\eta^2_p = 0.19$. These results suggest that searching for an incongruent target (866 ms, 90.5%) was slower and less accurate than for a congruent target (773 ms, 93.3%). The results also revealed a significant main effect of Distractor Congruency on the RTs, $F(1, 39)$
\[ F(1, 38) = 8.11, \ p = 0.007, \ \eta_p^2 = 0.17, \] suggesting that responses were slower when the distractors were incongruent (825 ms) than when they were congruent (814 ms). None of other main or interaction effects was significant, all \( F_s < 0.51, \ ps > 0.49. \)

Planned pairwise comparisons based on our research hypotheses revealed that responses were slower and less accurate in the incongruent target-incongruent distractor condition (872 ms, 90.1\%) than in the congruent target-incongruent distractor condition (777 ms, 93.1\%), \( t(39) = 12.03, \ p < 0.001, \) Cohen's \( d = 1.92, \) accuracy: \( t(39) = 2.53, \ p = 0.016, \) Cohen's \( d = 0.41. \) By contrast, responses were slightly slower in the incongruent target-incongruent distractor condition (872 ms, 90.1\%) than in the incongruent target-congruent distractor condition (860 ms, 91.0\%), \( t(39) = 1.99, \ p = 0.05, \) Cohen's \( d = 0.31, \) with comparable accuracy, \( t(39) = 0.81, \ p = 0.42. \)

**Neuroimaging results**

As shown in Figure 4, pairwise comparisons revealed greater activation in the incongruent target-incongruent distractor condition than in the congruent target-incongruent distractor condition in the right putamen, \( t(39) = 3.08, \ p = 0.004, \) Cohen's \( d = 0.49. \) By contrast, we found no significant differences between these two conditions in activations of the IPS, frontal eye fields, or the right ventral frontal cortex, all \( ts < 1.51, \ ps > 0.14. \) Moreover, none of the ROIs showed any significant
activation differences between the incongruent target-congruent distractor and incongruent target-incongruent distractor conditions, all $t_s < 1.25, p_s > 0.22$.

**Figure 4.** The ROI analyses revealed greater activation in the incongruent target-incongruent distractor condition (illustrated by the blue columns) than in the congruent target-incongruent distractor condition (illustrated by the red columns) within the right putamen.

As summarized in Table 2, the whole-brain Target Congruency × Distractor Congruency ANOVAs revealed a significant main effect of Target Congruency in the right precentral gyrus, suggesting greater activation for the incongruent targets than for the congruent targets. None of other main or interaction effects reached the significance level. That is, even though the behavioral results revealed a main effect of Distractor Congruency, we did not find any significant effects of Distractor Congruency in the neuroimaging data.
Table 2. Results of the whole-brain analyses in the fMRI study.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Hemisphere</th>
<th>MNI coordinates</th>
<th>Voxels</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>Main effect: Incongruent target &gt; congruent target</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>R</td>
<td>28</td>
<td>-24</td>
<td>56</td>
</tr>
<tr>
<td>Incongruent target-incongruent distractor &gt; congruent target-incongruent distractor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusiform</td>
<td>L</td>
<td>-14</td>
<td>-82</td>
<td>6</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>R</td>
<td>46</td>
<td>-6</td>
<td>34</td>
</tr>
</tbody>
</table>

Note: The voxel-level threshold was \( p < 0.01 \), and the cluster-extent threshold was \( k > 719 \).

As summarized in Table 2 and illustrated in Figure 5, planned pairwise comparisons based on our research hypotheses revealed greater activation in the incongruent target-incongruent distractor condition than in the congruent target-incongruent distractor condition, within the left fusiform and the right precentral gyrus. By contrast, no significant differences were found between the incongruent target-congruent distractor and incongruent target-incongruent distractor conditions.

Figure 5. Brain regions showing greater activations in the incongruent target-incongruent distractor condition than in the congruent target-incongruent
distractor condition (identified by the whole-brain analyses in the fMRI study), including the left fusiform and the right precentral gyrus.

2.3 Discussion

In the present study, we found the main effect of Target Congruency in visual search for food labels in our behavioral data, which replicated the findings of Velasco et al. (2015). Moreover, our behavioral results revealed a significant main effect of Distractor Congruency, suggesting that the participants responded more slowly when the distractors were incongruent than when they were congruent. However, we did not find a significant interaction term between Target Congruency and Distractor Congruency, implying that the color-flavor incongruency effect cannot be attributed to target-distractor similarity (Duncan & Humphreys, 1989).

Importantly, our ROI analyses revealed greater activation in the right putamen under the incongruent target-incongruent distractor condition in which the participants’ color expectations were violated than under the congruent target-incongruent distractor condition where their expectations were fulfilled. In contrast, the ROI analyses revealed no significant differences in the activation of the IPS, frontal eye fields, or the right ventral-frontal cortex between the incongruent target-incongruent distractor and incongruent target-congruent distractor conditions, which is inconsistent with the prediction of the primed attention-capture account. Therefore, our imaging results are more in line with the violation of expectation account of the color-flavor incongruency effect. That is, the participants might volitionally generate
expectation concerning the color of the target after seeing the flavor cue, and their visual search within the subsequent display may then be influenced by the violation of such expectation.

Our imaging results linked the right putamen to the violation of color expectation generated on the basis of flavor word, which is consistent with the findings of two meta-analyses separately conducted by D’Astolfo and Rief (2017) as well as Garrison et al. (2013). Moreover, our whole-brain analyses also linked the left fusiform and the right precentral gyrus to the color-flavor incongruency effect. Previous research has revealed that the V4 area, located in the fusiform cortex, is crucial for the perception and imagery of color in the normal population (Howard et al., 1998; Lueck et al., 1989), as well as for perceiving color in grapheme-color synesthesia (Nunn et al., 2002). Therefore, the results of the present study suggest that the participants did (at least partially) rely on color to guide their attention in visual searching for flavor labels. It is possible that the participants performed more processing of colors when the target was incongruent than when it was congruent, resulting in greater activation in the fusiform cortex in the former condition. In contrast, our findings regarding the right precentral gyrus are consistent with previous findings that stimulus-response incompatibility evokes greater activation in the right precentral gyrus compared to stimulus-response compatibility (Dassonville et al., 2001; Koski et al., 2005). Therefore, these results suggest the important role that precentral gyrus may play in conflict resolution.
Taken together, the results of the present study revealed that the functional basis of the color-flavor incongruency effect might involve the putamen, fusiform cortex, and precentral gyrus. In the following VBM study, we further examined whether individual differences in the color-flavor incongruency effect were associated with variability in the volume of gray matter in the areas linked to the violation of expectation or attention capture.

3. A VBM study of the color-flavor incongruency effect

3.1 Method

Participants

The VBM analyses were based on the data of 93 healthy participants (mean age = 22.73 ± 3.87 years, ranging from 18 to 37 years; 48 male and 45 female) from four previously acquired datasets: Sample 1 from the fMRI study reported above (n= 40), Sample 2 (n = 19, mean age = 24.3 ± 3.18 years; 9 males) and Sample 3 (n = 18, mean age = 25.0 ± 4.46 years; 7 male) from two piloting behavioral experiments conducted prior to the fMRI study, as well as Sample 4 (n = 16, mean age = 23.2 ± 4.79 years; 12 male) from one close replication of Velasco et al.’s (2015) first experiment. All participants performed the visual search task for flavor labels and underwent the T1-weighted high-resolution anatomical brain imaging on the exactly same MRI scanner with an identical imaging protocol.
Apparatus and stimuli

The behavioral experiments were conducted at the Psychology Department of Tsinghua University, with the same apparatus and stimuli used as those of the fMRI study except for the following differences. The stimuli were presented on 17-inch monitors with a resolution of 1024 × 768 pixels, at a refresh rate of 85 Hz. The participants viewed the displays when sitting at a distance of approximately 48 cm, and their responses were collected via keyboard presses. Specifically, as for the two piloting experiments that Samples 2 and 3 took part in, they were instructed to press the A, Z, K, or M keys on the keyboard when the target was presented in the top-left, bottom-left, top-right, or bottom-right quadrant of the screen, respectively. As for Sample 4, half of them were instructed to press the Z or M keys on the keyboard to indicate the presence or absence of the target, respectively, whereas the other half were instructed to press the M or Z keys to indicate the presence or absence of the target, respectively.

Design and procedure

The design and procedure for the two piloting experiments that Samples 2 and 3 took part in were the same as those of the fMRI study except for that each search display was presented for a maximum of 2500 ms for Sample 2. As for Sample 4, both target-present and target-absent trials were intermixed and presented in a random order, and the participants were instructed to press different keys to indicate whether
the target was present or absent. We still used a Target Congruency × Distractor Congruency within-participants design for the target-present trials; whereas the target-absent trials consisted of either congruent or incongruent stimuli. Each trial started with a 1000-ms fixation screen, followed by a 2000-ms screen presenting a flavor word. Next, another fixation screen was presented for 1000 ms, followed by a search display for a maximum of 1500 ms.

Imaging data preprocessing

Each image was first displayed in SPM8 (in MATLAB R2013b) to screen for artifacts or gross anatomical abnormalities. Then we performed T1-weighted anatomical registration through Exponentiated Lie (DARTEL) algebra for registration, normalization, and modulation (Ashburner, 2007), and segmented all T1 images into gray matter, white matter, and cerebrospinal fluid. We also applied a non-linear warping of the gray matter images to the DARTEL gray matter template in the MNI space at a voxel size of $1.5 \times 1.5 \times 1.5 \text{ mm}^3$. After that, we smoothed the modulated gray matter images using an 8-mm full-width half-maximum Gaussian kernel to reduce spatial noise.

VBM analyses

We performed the whole-brain analysis by entering all the resulting gray images into a multiple linear regression analysis implemented in SPM8 to determine the relationship between the regional gray matter volume ($rGMV$) and the magnitude of
the color-flavor incongruency effect. That is, similar to what we did for the fMRI study, we focused on our analyses on the data of the congruent target-incongruent distractor and incongruent target-incongruent distractor conditions. To rule out the influence of potential confounding factors, we also included the total intracranial volume, age, sex, and samples as covariant variables in this analysis. Statistical maps were thresholded at $p < 0.01$, and the issues of multiple comparisons were addressed with a cluster extent threshold of 2042 voxels (equivalent to corrected $p < 0.05$), estimated by the 3dClustSim program of AFNI software using the ACF approach (ACF values: 0.43, 4.20, 17.86; 5000 iterations).

After that, we extracted gray matter values from a 10 mm sphere centered around the maximally correlated voxels of regions (i.e., the peak coordinates of the clusters) identified by the whole-brain analyses in the present study, and calculated the partial correlations between the rGMV of these regions and the color-flavor incongruency effect, with covariant variables controlled for.

3.2 Results

*Behavioral results*

In the present study, the participants showed a high level of accuracy of 93.47% overall. We excluded RTs which were three standard deviations shorter or longer than the group means from the following data analyses, resulting in 4.94% of the data
being discarded. The mean RTs based on correct trials and accuracy data for each condition are shown in Figure 6.

**Figure 6.** Mean RTs (in ms) and accuracy for each dataset included in the VBM study. Note that * denotes $p < .05$, *** denotes $p < 0.001$, and the error bars show the standard errors of the means.

We performed 2 (Target Congruency: congruent or incongruent) $\times$ 4 (Sample: Sample 1, 2, 3, or 4) mixed-design ANOVAs on the RTs, with Target Congruency
being the within-participants factor and Sample being the between-participants factor.

The results revealed a significant main effect of Target Congruency on the RTs, $F(1, 89) = 162.71$, $p < 0.001$, $\eta_p^2 = 0.65$, and on the accuracy data, $F(1, 89) = 15.86$, $p < 0.001$, $\eta_p^2 = 0.15$. These results suggest that responses to the incongruent targets (923 ms, 92.8%) were slower and less accurate than to the congruent targets (845 ms, 95.7%). The results also revealed significant main effects of Sample on the RTs, $F(3, 89) = 7.16$, $p < 0.001$, $\eta_p^2 = 0.19$, and on the accuracy, $F(3, 89) = 4.15$, $p = 0.008$, $\eta_p^2 = 0.12$. Pairwise comparisons with Bonferroni corrections revealed that participants from Sample 1 made less accurate responses than those from Sample 2, $t(57) = 2.83$, $p = 0.01$, Cohen’s $d = 0.57$, and made faster responses than those from Sample 4, $t(54) = 5.21$, $p < 0.001$, Cohen’s $d = 1.16$; whereas none of any other pairwise comparison on the RT or accuracy data reached the significant level, all $t$s $< 2.32$, $p$s $> 0.12$.

Importantly, the interaction term between Target Congruency and Sample was not significant on the RTs, $F(3, 89) = 1.52$, $p = 0.22$, nor on the accuracy data, $F(3, 89) = 1.80$, $p = 0.15$. Planned pairwise comparisons revealed that Samples 1, 2, 3, and 4 showed an average of 95 ms, 77 ms, 80 ms, and 61 ms color-flavor incongruency effect, respectively, all $t$s $> 4.53$, $p$s $< 0.001$. These color-flavor incongruency effects were comparable to each other, all $t$s $< 2.42$, $p$s $> 0.25$. Therefore, we calculated the magnitude of the color-flavor incongruency effect for each participant by subtracting the RTs of the congruent target-incongruent distractor condition from those of the incongruent target-incongruent distractor condition.
Voxel-based correlation analyses between the color-flavor incongruency effect and rGMV

As summarized in Table 3, the multiple linear regression analysis on the whole brain revealed that the magnitude of the color-flavor incongruency effect was positively correlated with the rGMV in a cluster of voxels located in the right putamen and caudate (which peaked in putamen), and with the rGMV in a cluster of voxels located in the left putamen, caudate, and thalamus (which also peaked in putamen).

Table 3. Partial correlations between the color-flavor incongruency effect and rGMV in the VBM study (N=93).

<table>
<thead>
<tr>
<th>Regions</th>
<th>Hemisphere</th>
<th>MNI coordinates</th>
<th>Voxel</th>
<th>t</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putamen/caudate</td>
<td>R</td>
<td>17 21 -9</td>
<td>2227</td>
<td>5.17</td>
<td>0.25*</td>
</tr>
<tr>
<td>Putamen/caudate/thalamus</td>
<td>L</td>
<td>-23 15 -11</td>
<td>2242</td>
<td>5.38</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Note: The voxel-level threshold was \( p < 0.01 \), and the cluster-extent threshold was \( k > 2042 \).

As summarized in Table 3 and illustrated in Figure 7, the partial correlation analyses also revealed that the rGMV in the right putamen was positively correlated with the magnitude of the color-flavor incongruency effect. Therefore, both analyses consistently revealed that the participants having greater volume of gray matter in the right putamen showed greater color-flavor incongruency effects.
Figure 7. The rGMV in the right putamen was positively correlated with the magnitude of the color-flavor incongruency effect. Note that Y-axis shows the adjusted color-flavor incongruency effect with covariates being regressed out, and * denotes $p < .05$.

3.3 Discussion

In the present study, we combined the behavioral and structural imaging data across four previously acquired datasets. As for the behavioral data, we did observe some significant differences in the overall RTs or accuracy data between some of the samples. These differences might be attributed to the participants’ speed-accuracy tradeoff elicited by the varied time parameters being used, such as the maximum allowable response time (1500 ms for Sample 1 vs. 2500 ms for Sample 2) or the presentation duration of the flavor word (1000 ms for Sample 1 vs. 2000 ms for Sample 4). However, these four samples showed comparable color-flavor incongruency effects. This result suggests that the color-flavor incongruency effect is robust, regardless of different time parameters, inclusion of the target-absent trials (as
for Sample 4), or task demand (indicating target absence/presence for Sample 4 vs. indicating target location for the other three samples).

Our VBM analyses revealed a significant positive correlation between the magnitude of the color-flavor incongruency effect and the rGMV in the right putamen. These results therefore suggest that those individuals having a greater volume of gray matter in the right putamen showed larger color-flavor incongruency effects. This result not only revealed the anatomical basis of individual differences in the color-flavor incongruency effect but also suggest an association between the right putamen and the color-flavor incongruency effect. This finding is consistent with what we found in the fMRI study (i.e., greater activation was observed in the incongruent target condition than in the congruent condition within the right putamen).

Furthermore, our whole-brain analyses also linked the color-flavor incongruency effect to the caudate and thalamus, two areas that have been previously considered to reflect the processing of violation of expectation (D’Astolfo & Rief, 2017; Garrison et al., 2013).

4. General Discussion

In summary, the behavioral results of both the fMRI and VBM studies revealed that the participants were slower to find the target when its color was incongruent with their color-flavor associations than when it was congruent, which is indicative of a color-flavor incongruency effect in visual search for food labels. The functional
imaging results revealed greater activation to the incongruent targets than to the congruent targets in the right putamen, which is associated with the violation of expectation (D’Astolfo and Rief, 2017; Garrison et al., 2013). Collectively, these results suggest that this color-flavor incongruency effect might be attributed to volitional control based on the participants’ color expectations. When the target was incongruent with the participants’ color expectations, such violation of expectation might lead a less efficient search than that in the congruent condition, resulting in the color-flavor incongruency effect.

The flavor word used to cue the visual search for a certain food label in both the present study and Velasco et al.’s (2015) study might be the key reason why the violation of expectation plays an important role. That is, the Chinese character for “flavor” was used for packaging printed in Chinese, such as “tomato-flavor,” “cucumber-flavor,” “chicken-flavor,” or “BBQ-flavor” (see Figure 2a for illustrations of these flavor labels); however, only the “object” (such as “tomato,” “cucumber,” “chicken,” or “BBQ”) was used for flavor labels for potato-chips packaging printed in English or Spanish (Piqueras-Fiszman et al., 2012; Velasco et al., 2014). Therefore, it is possible that the participants in the present study and Velasco et al.’s (2015) study holistically processed the three-character word such as “tomato-flavor,” instead of merely extracting two characters expressing the object and processing it. In other words, the fashion by which flavor labels are printed in Chinese might emphasize the
flavor as a unit to process and elicit expectations regarding the color of the target with this flavor label.

Moreover, it should also be noted that Velasco et al. (2015) also proposed the behavioral effect in visual search for flavor labels as a Stroop-like effect. A classical Stroop effect refers to Stroop’s (1935) demonstration that people were slower and less accurate to report the ink color of a word expressing color when they were incongruent with each other (e.g., a word “yellow” printed in red ink), compared to when they were congruent (e.g., a word “yellow” printed in yellow ink). Previous studies have linked the Stroop effect to the anterior cingulate cortex and dorsolateral prefrontal cortex (Adleman et al., 2002; Banich, Duncan, Brett, & Lawrence, 2000; Milham, Banich, Claus, & Cohen, 2003; Mitchell, 2005; Pardo, Pardo, Janer, &Raichle, 1990; Peterson et al., 1999). Nevertheless, the results of our fMRI and VBM studies did not reveal any associations between these two areas and the color-flavor incongruency effect.

Admittedly, there are also several limitations to the present study. For one, each search display only consisted of three images of packaging in our experiments, whereas people often encounter more packaging at a time in everyday life. Therefore, it will be interesting to increase the set size of search displays in future studies to examine the neural basis of the color-flavor incongruency effect. For another future study, it will also be interesting to explore the implications of the color-flavor
incongruency effect in everyday scenarios such as product selection or evaluation, which is currently being pursued in our laboratory (Huang & Wan, submitted).

In conclusion, the findings of the present study have associated the right putamen with the color-flavor incongruency effect and, therefore, link this behavioral effect to the violation of expectation. Hence, these findings provide insights into the brain areas involved in the violation of expectation.
Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 71872097) awarded to Xiaoang Wan. The authors would like to thank Yuxuan Qi, Weiwei Zhang, and Yihang Ouyang for their technical assistance.

Conflict of Interest Statement

None.
References


Highlights

- Searching for a package was slower when its color was incongruent with flavor label.
- This behavioral effect was associated with greater activation in the right putamen.
- This effect was positively correlated to the volume of gray matter in the right putamen.
- The color-flavor incongruency effect was attributed to violation of color expectation.
- Putamen is linked to the processing of events in violation of expectation.