Electrophysiological evidence for effects of color knowledge in object recognition

Aitao Lu, Guiping Xu, Hua Jin, Lei Mo, Jijia Zhang, John X. Zhang

Abstract

Knowledge about the typical colors associated with familiar everyday objects (i.e., strawberries are red) is well-known to be represented in the conceptual semantic system. Evidence that such knowledge may also play a role in early perceptual processes for object recognition is scant. In the present ERP study, participants viewed a list of object pictures and detected infrequent stimulus repetitions. Results show that shortly after stimulus onset, ERP components indexing early perceptual processes, including N1, P2, and N2, differentiated between objects in their appropriate or congruent color from these objects in an inappropriate or incongruent color. Such congruence effect also occurred in N3 associated with semantic processing of pictures but not in N4 for domain-general semantic processing. Our results demonstrate a clear effect of color knowledge in early object recognition stages and support the following proposal—color as a surface property is stored in a multiple-memory system where pre-semantic perceptual and semantic conceptual representations interact during object recognition.

Article history:
Received 17 July 2009
Received in revised form 12 December 2009
Accepted 15 December 2009

Keywords:
Color knowledge
Object recognition
Surface property
Shape
Semantic knowledge

The brain stores knowledge about the particular colors typically associated with familiar everyday objects, presumably acquired through repeated experiences with them throughout one’s life [4]. Research has shown that such object color knowledge (e.g., knowing a strawberry is red) can be verbalized and represented in the conceptual semantic system. Lesion evidence from neuropsychological studies has documented dissociations between color knowledge and color perception, between color knowledge and other aspects of object knowledge. For example, impairment to lingual gyrus often results in achromatopsia sparing color knowledge, but impairment to ventral temporal cortex usually leads to color agnosia leaving color vision intact [3,22,37]. However, it is unclear whether such object color knowledge is also stored in pre-semantic perceptual representations critical to object recognition. Edge-based theories propose that object recognition is based primarily on shape properties but not much on surface properties, such as color or texture [11,29,35,39,42]. By these theories, color effects in object recognition reflect facilitation at the semantic level for concept identification [16,27]. Surface-plus-edge-based theories, however, allow for object representations to include information about surface properties and claim that color can facilitate object recognition at the perceptual level [15,46]. That is, color, as a surface property, is stored in a multiple-memory system with both pre-semantic sensory and semantic conceptual representations [31,32,44]. Neuropsychological evidence also supports a finer division of object color knowledge into a verbal format and a visual format that are closely bound normally but can be dissociated in patients [1,3].

Using brain imaging, Zeki and Marini [47] obtained some initial evidence in the normal population for a connection between color knowledge and object recognition processes. They asked participants to view normally colored objects (e.g., red strawberries), abnormally colored objects (e.g., blue strawberries), and gray-scaled Mondrian displays. They found that appropriately and inappropriately colored objects activated different pathways and demonstrated an effect of color knowledge on object recognition [47]. In our view, the Zeki and Marini design, if combined with the event-related potential (ERP) technique capable of distinguishing early and late perceptual processes, would offer an excellent rationale to examine whether color knowledge exists in pre-semantic perceptual representations. There has been so far only one study along this direction. Proverbio and colleagues asked participants to detect a target in a list of colored objects by paying attention to either the color or the shape dimension. For the attend-to-color condition, the N2 component was found to be significantly greater when an object was in its typical color, relative to a non-typical color [29]. Although such effects were absent in the attend-to-shape condition, this study shows clear effects of object color knowledge at a very early perceptual stage.

While the task in Proverbio et al. demands selective attention to either the color or shape dimension, everyday object recognition typically relies on identification of an object as a whole without
explicitly attending to its color. It is therefore unclear whether their
d results would differ in more ecological situations.

In the present study, we followed up on Proverbio et al. to see
if we could optimize the conditions and observe effects of object
color knowledge when the task was primarily based on processing
an object as a whole. It is hoped that the task would simulate a more
natural scenario where attention to color is neither encouraged nor
discouraged. In this task, common everyday objects either in their
canonical colors or in some incongruent colors were presented one
by one. Participants were to detect infrequent repetitions of some
filler items.

To avoid confounds from motor response execution, we focused
on ERP waveforms elicited by non-filler objects. Using the same
critical manipulation as in the Proverbio et al. study, we com-
pared ERP components for objects presented in their appropriate
color and objects in some inappropriate color. These components
include N1, P2, and N2 that have been widely demonstrated
to reflect early perceptual processes. N1 is the first negative
component elicited 80–130 ms post-visual stimulus onset [24].
It involves multiple generators in temporal, frontal, cingulate,
and parietal regions [25,40] and has been shown to be sensi-
tive to faces and view-invariant object discrimination [17,33,43].
N2 is a widespread negative potential peaking between 200 and
350 ms post-stimulus. Traditional research on N2 mostly focuses
on its role in stimulus classification [28,36] and stimulus deviance
detection [6]. P2 is evoked about 200 ms post-stimulus and its
amplitude is modified by attention and working memory manip-
ulations [21,23,30]. N2 has been found mainly in priming tasks
and interpreted as a neural correlation of repetition suppression
[13].

<table>
<thead>
<tr>
<th>Sample test pictures</th>
<th>Congruent color</th>
<th>Incongruent color</th>
<th>Achromatic gray color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td><img src="image" alt="Apple" /></td>
<td><img src="image" alt="Incongruent Apple" /></td>
<td><img src="image" alt="Achromatic Apple" /></td>
</tr>
<tr>
<td>Banana</td>
<td><img src="image" alt="Banana" /></td>
<td><img src="image" alt="Incongruent Banana" /></td>
<td><img src="image" alt="Achromatic Banana" /></td>
</tr>
<tr>
<td>Carrot</td>
<td><img src="image" alt="Carrot" /></td>
<td><img src="image" alt="Incongruent Carrot" /></td>
<td><img src="image" alt="Achromatic Carrot" /></td>
</tr>
<tr>
<td>Broccoli</td>
<td><img src="image" alt="Broccoli" /></td>
<td><img src="image" alt="Incongruent Broccoli" /></td>
<td><img src="image" alt="Achromatic Broccoli" /></td>
</tr>
</tbody>
</table>

| Sample filler pictures       | ![Besom](image) | ![Folder](image) | ![Tyre](image) | ![Pan](image) |

Fig. 1. Sample test and filler pictures used in the experiment.
Other than these early components, we expected to see color knowledge effects in later ERP components such as N3 and N4. N3 has been generally recognized to be sensitive to the semantic rather than the physical properties of picture stimuli [8]. N4 is a component particularly relevant to semantic processing [9,18,19,38,41], peaking around 400 ms post-stimulus in tasks involving semantic or conceptual processing.

Although both N3 and N4 are more negative-going for semantic incongruence [7,12,20,45], they seem to reflect different aspects of object recognition. N3 has been considered to reflect picture-specific semantic processes. However, N4 has been observed in studies concerning a variety of meaningful representations [2,5,14,26,46] and considered to reflect more general semantic processes. Given object color knowledge is represented in semantic memory, manipulation of an object's shape-color relationship may also modulate these two components.

Fifteen right-handed healthy undergraduates (seven males, mean age = 23.2 years, SE = 2.1) participated in the experiment. All had normal or corrected-to-normal vision and normal color vision. Informed consent was obtained following an IRB approved research protocol.

Ninety-six pictures of everyday objects were selected from Photodisc collection (Photodisc Inc., Seattle, WA), half for test items and half for filler items. Test pictures represent color-diagnostic natural objects that tend to have one typical color. Each was edited to create three versions. An object (e.g., apple) was either in its typical color for the congruent version (e.g., red apple), in a non-typical color for the incongruent version (e.g., blue apple), or in a gray achromatic control version (e.g., gray apple). Mean luminance was matched for different versions. Filler pictures represent man-made objects (e.g., besom), randomly colored in one of six colors (purple, blue, red, yellow, green, and gray) to reduce explicit attention to color object associations in the test items. Fig. 1 shows some sample test and filler pictures. All pictures fitted in a 250 × 250 pixel area (6.2° × 6.2° in visual angle) at screen center.

A separate group of 20 participants rated the congruent and incongruent versions of each test picture on whether the color matches the object in real life on a Likert scale from 1 (poorly...
matched) to 7 (well-matched). Mean ratings for the incongruent version were significantly lower than the congruent version (2.02 vs. 6.46, \( p < 0.001 \)). They also rated image clarity of all test pictures on a Likert scale from 1 (poor quality) to 7 (good quality). All pictures scored more than 5 and the mean ratings did not differ across conditions (\( p > 0.5 \)).

Participants were seated in a dimly lit and sound-attenuated room and instructed to refrain from blinking, eye movements, and swallowing during stimulus presentation. Pictures were presented in a white background. A trial began with a fixation lasting 300–500 ms. After a 500 ms blank screen, a picture was shown for 1000 ms followed by a 500 ms blank interval before the next trial started. Participants were required to detect an immediate picture repetition. For example, if an apple picture was preceded by the same picture, regardless of their colors, they should press the space bar as quickly as possible. If the two neighboring pictures did not match in shape, no response was required. The repetition occurred only for some of the filler pictures but never for the test pictures. Each block contained 128 trials, including 48 test trials and 48 filler trials, and 24 repetitions trials with filler pictures. All 96 pictures were different from each other. The test pictures were divided into three sets of 16 pictures, in their congruent, incongruent, and gray versions, respectively. Each participant had one block of 15 practice trials and three blocks of test trials. The test block orders were counterbalanced across-subjects.

EEGs were continuously recorded from 64 non-polarizable Ag/AgCl sintered electrodes mounted in an elastic cap by the 10–20 system, referenced off-line to mean of the two mastoids. Vertical eye movements were monitored via supra- to sub-ornital bipolar montage. EEG and EOG data were amplified with a BrainAmp MR plus EEG amplifier. Electrode impedances were below 3 kΩ for EEG recording and below 5 kΩ for EOG recording. The EEG and EOG signals were digitized on-line (sampling rate = 1000 Hz).

Eye movement artifacts were removed using regression-based weighting coefficients. Epochs containing artifacts exceeding ±80 µV were automatically excluded. Average waveforms were calculated off-line for all trial conditions in the −200 to 800 ms epoch, relative to picture onset (Fig. 2).

Participants were highly attentive to the task as they performed at ceiling level for repetition detection (mean response time = 609 ms, SD = 65). Grand average waveforms show five time windows of interest: three early components: N1 (80–150 ms), P2 (150–200 ms), N2 (200–275 ms), and two later components: N3 (275–375 ms) and N4 (375–475 ms).

Three-way 3 × 3 × 4 repeated-measure ANOVAs were conducted on mean amplitudes of these components in representative electrodes in frontal-central-parietal regions where modulations on ERPs by trial condition were mostly seen. The three factors were trial type (congruent, incongruent, gray versions), laterality (left, right hemisphere, midline), and lobe (frontal: F3, Fz, F4; frontal-central: FC3, FCz, FC4; central: C3, Cz, C4; central-parietal: CP3, CPz, CP4).

For this component occurring 80–150 ms post-stimulus, ANOVA revealed a main effect of lobe, \( F(3, 42) = 4.19, \) \( MSE = 27.41, \) \( p < 0.05 \), showing more negativity in frontal-central (−2.58 vs. 1.58 µV, \( p < 0.05 \)) and central regions (−2.38 vs. −1.58 µV, \( p < 0.05 \)) than central-parietal regions. There was also a main effect of trial type, \( F(2, 28) = 5.59, \) \( MSE = 50.69, \) \( p < 0.01 \). The achronamic and the incongruent trials were not different from each other (\( p > 0.5 \)), but both were more negative than the congruent trials (−2.57 vs. −1.63 µV, \( p = 0.005 \), −2.53 vs. −1.63 µV, \( p < 0.05 \)). There was no significant interaction between trial type and other factors (\( F(5, 1) \)).

For P2 at 150–200 ms and N2 at 200–275 ms time windows, ANOVAs showed a main effect of trial type (\( F(2, 28) = 3.95, \) \( MSE = 67.04, \) \( p < 0.05 \); \( F(2, 28) = 5.57, \) \( MSE = 72.51, \) \( p < 0.05 \)). For both components, the congruent condition showed more positivity than the gray (P2: 1.96 vs. 0.83 µV, \( p < 0.05 \); 1.96 vs. 1 µV, \( p < 0.05 \); N2: 0.89 vs. 0.06 µV, \( p < 0.05 \); 0.89 vs. −0.36 µV, \( p = 0.005 \)) and there were no differences between the latter two (\( ps > 0.1 \)). There was no significant interaction between trial type and other factors (\( F(5, 1) \)).

For this component occurring at the 275–375 ms time window, ANOVA showed main effects of lobe, \( F(3, 42) = 25.29, MSE = 212.26, \) \( p < 0.001 \), and trial type, \( F(2, 28) = 11.06, MSE = 130.56, \) \( p < 0.001 \). There were no interaction effects involving trial type (\( p > 0.1 \)).

The congruent condition was more positive than the incongruent and the gray conditions (2.34 vs. 1.43 µV, \( p < 0.05 \); 2.34 vs. 0.64 µV, \( p < 0.001 \)). The incongruent condition was more positive than the gray condition (1.43 vs. 0.64 µV, \( p = 0.07 \)).

ANOVA on N4 at the 375–475 ms time window revealed a main effect of lobe, \( F(3, 42) = 15.7, MSE = 127.97, \) \( p < 0.001 \), showing progressive and significant amplitude increases from frontal, frontal-central, central regions, to central-parietal regions (0.37, 1.03, 1.78, 2.63 µV, for all pair-wise comparisons, \( p < 0.01 \)).

The main effect of trial type was significant, \( F(2, 28) = 4.89, MSE = 101.55, \) \( p < 0.05 \). The congruent condition was more positive than the gray condition (2.21 vs. 0.71 µV, \( p = 0.005 \)), but there were no differences between the congruent and incongruent conditions (2.21 vs. 1.43 µV, \( p > 0.1 \)).

In the present study, we examined whether knowledge about the typical colors associated with familiar everyday objects plays a role in early perceptual processes during object recognition. We found that manipulation of an object’s color modulated ERP components indexing early perceptual processing. Briefly, shortly after the onset of a familiar object, the brain differentiated between an object in its appropriate or congruent color from an object in an inappropriate or incongruent color, as indicated by significant positive shifts of N1, P2, and N2 in the former relative to the latter.

As the same set of shapes and colors were used fully counterbalanced across conditions, the observed ERP congruence effects should be attributed only to the nature of the color-shape associations but not to any sensory or physical characteristics of the stimuli. These results thus indicate color is represented in perceptual memory of objects and automatically activated during object recognition.

One result is that the gray condition essentially overlapped with the incongruent condition in early components. This seems to indicate that gray is processed like an incongruent color. The gray condition may also be considered as a non-color neutral control to assess the benefits or costs (i.e., facilitation and interference) color information brings to object recognition. The ERP responses in the gray condition differed from the congruent condition but not from the incongruent condition. This suggests that the congruence effect we observed reflects primarily a benefit but not a cost. That is, color information would facilitate object recognition when it matches an object’s typical color but it would not impede object recognition when it does not. This supports the usual characterization of color effects in object recognition as facilitation [24].

Note that Proverbio et al. [29] failed to find color knowledge effects in early perceptual processing for their attend-to-shape condition. This is puzzling as one would expect shape, being most salient in object recognition, to automatically activate an object’s associated color information. We suspect that, firstly, their explicit instruction to ignore color in this condition may have suppressed the salient in object recognition, to automatically activate an object’s associated color information. We suspect that, firstly, their explicit instruction to ignore color in this condition may have suppressed the color dimension. When no such suppression of the color dimension was required in the present study, the color effects were demonstrated. Secondly, early ERP components are known to be sensitive to stimulus repetition. With nearly two hundred repetitions for each item in a small set of 16 objects, these early components in the Proverbio et al. study may be saturated and not at an optimal level to reflect effects from manipulations of object color associations. In our study, each object was repeated only three times, the minimal allowed by the design. This may have helped avoid such
habituation effects. Thirdly, while they found more negative-going N2 for congruent objects than incongruent ones, we found more positive-going early components including N2 for the same comparison that is more consistent with typical results in the literature [10,34]. Proverbio et al. used 16 objects, each in eight different colors, one being the typical and the other seven non-typical so the congruent: incongruent trial number ratio was 1:7. In comparison, the ratio was 1:1 in our design (excluding the gray pictures as non-color pictures). Given ERP responses are sensitive to stimulus probabilities, this ratio factor may have caused some of the different results across the two studies.

Briefly, although more research is needed to understand the discrepancies between our results and the Proverbio et al. results, the present study does demonstrate a clear effect of object color knowledge in a more ecological situation of object recognition. This strengthens the Proverbio et al. study and supports the notion that object color knowledge is stored in perceptual representations.

A previous distinction between object identification/classification and memory recognition of objects suggests that the former relies more on an achromatic description system and the latter on a surface-based episodic memory system storing color information [44]. Our task involves a memory recognition component as participants maintained an object across a 1.5 s delay interval for matching with the incoming next object. Therefore, color may have been encoded and produced the congruence effects we observed. The implication is that color object knowledge is more likely to play a role in tasks involving recognition memory of objects, compared with simple identification or classification of isolated objects.

Finally, our results that object color knowledge also affected N3 supports the notion that such knowledge is represented in the semantic memory about pictures. During the N3 time window, the gray color condition started to diverge from the incongruent condition (though not significantly) and was more negative than the two color conditions. This possibly reflects a processing distinction between color and non-color objects.

For N4, unlike N3 and earlier components, the difference between congruent and incongruent conditions disappeared, suggesting that the activation of color knowledge had been complete. Though the difference between the congruent color and the gray color was still significant (2.21 vs. 0.71 μV), the effect was smaller relative to the effect in N3 (2.34 vs. 0.64 μV). The N4 pattern is consistent with the general understanding that N4 reflects modality-non-specific semantic representations [5,14,26,46].

References


