The effect of category learning on attentional modulation of visual cortex

Jonathan R. Folstein*, Kelly Fuller, Dorothy Howard, Thomas DePatie

Florida State University, United States

ARTICLE INFO

ABSTRACT

Learning about visual object categories causes changes in the way we perceive those objects. One likely mechanism by which this occurs is the application of attention to potentially relevant objects. Here we test the hypothesis that category membership influences the allocation of attention, allowing attention to be applied not only to object features, but to entire categories. Participants briefly learned to categorize a set of novel cartoon animals after which EEG was recorded while participants distinguished between a target and non-target category. A second identical EEG session was conducted after two sessions of categorization practice. The category structure and task design allowed parametric manipulation of number of target features while holding feature frequency and category membership constant. We found no evidence that category membership influenced attentional selection: a postero-lateral negative component, labeled the selection negativity/N250, increased over time and was sensitive to number of target features, not target categories. In contrast, the right hemisphere N170 was not sensitive to target features. The P300 appeared sensitive to category in the first session, but showed a graded sensitivity to number of target features in the second session, possibly suggesting a transition from rule-based to similarity based categorization.

1. Introduction

Attentional selection permeates most aspects of everyday cognition. In conversations, we select the voice we are listening to while filtering out others. When driving, we should select relevant road signs to tell us where we are and what we are allowed to do, while filtering out irrelevant stimuli such as people on the sidewalk. When our spatial attention lands on an object, we must decide whether it is in fact what we are looking for by evaluating its features. For instance, if we see a red shape by the road, we might also check to see if it is an octagon with the word "STOP" on it. The process of "checking" candidate targets is commonly thought to include acts of attention to the checked features in which features consistent with a relevant stimulus are selected and features inconsistent with a relevant stimulus are not (Hillyard and Anllo-Vento, 1998). Which features are selected and when is likely to vary from object to object and possibly task to task. It is quite plausible that one might be on the lookout specifically for stop signs, but stop signs are just one of a large and perceptually variable category of stimuli with importance for driving. Road signs alone are quite diverse. In American streets, they can be red, white, or yellow and might include multiple icons and symbols (e.g. child crossing, cow crossing, no u-turn, no-left turn, etc.). Under some circumstances all might be of equal importance as targets, resulting in the need to select them as a category rather than identify a single target with a single conjunction of features,

like a stop sign. How might attentional selection operate under these circumstances?

Consider the case of a pedestrian walking down the sidewalk wearing a t-shirt depicting a road sign, perhaps a "no u-turn" sign, or perhaps even a single component of such a sign, like a large "U". Clearly this is an irrelevant stimulus to a driver because it is not a real road sign – i.e. not in the road sign category. But would the picture on the t-shirt be attentionally selected nevertheless and only rejected at a later stage of processing, or might the visual system allow rapid rejection of the pedestrian prior to any feature selection? In other words, can knowledge of category membership modulate attention, engaging it in some cases and counteracting it others?

Some previous work suggests that attentional selection, which is indexed by midrange event-related potential (ERP) components with onsets of about 160 ms or later, occurs prior to categorization and operates on features. One ERP component, the selection negativity, is elicited by irrelevant stimuli sharing features with a target stimulus or a stimulus requiring further attentional monitoring (Harter and Aine, 1984; Hillyard et al., 1998). One can think of a target stimulus as a category with a single member, so attentional selection of non-targets that share features with a target would suggest that selection occurred at the feature level and was not counteracted by membership in the non-target category. Whereas these studies support feature based selection, they might not tell the whole story. First, they often use simple

---

psychophysical stimuli, such as conjunctions of color and simple shape or line orientation (Anllo-Vento and Hillyard, 1996; Kenemans et al., 1993; Martınez et al., 2001; McGinnis and Keil, 2011; Smid et al., 1999), and second they do not study how attention effects change with learning. Indeed, several studies suggest that, for well known natural objects, category information is available quite early, prior to the onset of the selection negativity. Rapidly presented animal or car targets have been shown to elicit EEG effects relative to non-targets within 150 ms (VanRullen and Thorpe, 2001) and object representations within orbito-frontal cortex are accessed by 130 ms (Bar et al., 2006), however the designs in these studies cannot rule out effects driven by recognition of category relevant features as opposed to fully formed category or exemplar representations.

In contrast to studies using natural objects as stimuli, Freedman and colleagues used morphed stimulus sets that allowed simultaneous assessment of neural sensitivity to stimulus features and categories in monkeys (Freedman and Assad, 2006; Freedman et al., 2003). In these studies, monkeys were trained to categorize morphed objects or psychophysical stimuli that varied in direction of motion. The design allowed comparison of stimuli that either belonged to same category or not even though perceptual similarity was held constant. Category sensitive neurons had similar firing rates to stimuli in the same category and different firing rates for stimuli in different categories. Feature sensitive neurons had firing rates tuned to perceptual features, regardless of category membership. Most category sensitive neurons were found in prefrontal and parietal cortex with latencies of 170 ms and 100 ms respectively. These studies did not assess attentional selection, but the latency of category sensitivity was well within the range required to plausibly allow attentional selection of categories and the most category sensitive cells were in prefrontal and parietal cortex, not far from attentional areas.

The studies of Freedman et al. show early sensitivity to category in higher areas that could plausibly control attention, but the unit of attentional selection is also likely to depend on the underlying perceptual representations upon which attention acts. Importantly, the nature of perceptual representations changes with learning, which increases neural and perceptual sensitivity to learned features, parts, and dimensions (Goldstone and Steyvers, 2001; Schyns and Rodet, 1997; Sigman et al., 2005) and the degree to which conjunctions of features are represented as unitized wholes (Baker et al., 2002; Gauthier and Tarr, 1997; Goldstone, 2000; Shiffrin and Lightfoot, 1997). There is also some evidence for perceptual representation of categories. In the phenomenon known as categorical perception, objects in the same category become perceptually more similar to each other than to members of other categories. Categorical perception has been observed most famously in phonemic categories (Harnad, 1987) but has also been observed for newly learned object categories both behaviorally (Goldstone, 1994; Livingston and Andrews, 1995; Wallraven et al., 2013) and neurally in perceptual cortex (Folstein et al., 2013; Mirabella et al., 2007).

If we assume that attentional selection of target objects is mediated by a perceptual template for relevant objects, both unitized perception of feature conjunctions and categorical perception could plausibly facilitate attentional selection of categories. To the degree that objects in a category are more similar to each other than to objects in other categories, they will more easily match a common template, or small number of templates. In contrast to categorical perception, unitized perception of feature conjunctions allows each exemplar to become more dissimilar from other exemplars that share some of its features, regardless of category membership (e.g. Shiffrin et al., 1997). However, to the degree that perceptual templates could be held in mind for each exemplar, holistically perceived exemplars of target categories should elicit similar attentional selection, which should be greater than selection of non-target categories.

On the other hand, if learning enhances the perception of individual object parts or features, attentional selection could become more

sensitive to object features, where what is meant by "feature" will depend on the perceptual unit that is relevant for categorization. For instance, if color is diagnostic for category membership, perceptual learning will occur along the color dimension, enhancing attentional selection of color (Goldstone, 1994). If more complex units such as object parts are relevant, attentional selection will become more sensitive to object parts (Sigala and Logothetis, 2002). This account is consistent with long standing mathematical models of category learning, which account for categorization behavior giving greater weight to category relevant perceptual dimensions (Maddox, 2001; Nosofsky, 1986) and with eye movement and priming studies finding that attention is directed to category-relevant perceptual features (Baruch et al., 2015; Blair et al., 2009).

The brief review above suggests that attentional selection of categories could be plausible, but could also depend on the type of perceptual units enhanced or created by a given category learning task. Furthermore, studies of detection of single targets seem to suggest attentional selection largely at the feature level. Our goal in the current study was therefore to contrast attentional selection of categories with attentional selection of features. Past studies of attentional selection have used event related potentials. We now review two ERP components known to be sensitive to attentional selection, the selection negativity and the N250.

1.1. ERP components sensitive to attention: selection negativity and N250

Event related potentials have long been employed to study the neural effects of attention and have often proven useful in resolving stages of processing affected by various types of attention, including object-based attention, feature-based attention, and spatial attention (Hillyard et al., 1998; Woodman and Luck, 1999; Zhang and Luck, 2009). Particularly relevant to the current experiment is the selection negativity (SN), an ERP component sensitive to the presence of task-relevant stimulus features. Some studies suggest that the SN can be divided into two epochs with distinct scalp distributions: an early epoch from about 160–200 ms with a lateralized scalp distribution when attended stimuli are presented laterally, and a later bilateral epoch from about 230–300 ms. The latency of the SN is variable and appears to be stimulus and task dependent (Anllo-Vento and Hillyard, 1996: 225–300 ms (late SN); Anllo-Vento et al., 1998: 250–350 ms (late SN); Smid et al., 1999: 204–388 ms (Shape Easy condition)). Two major paradigms are used to elicit the SN: rare target detection paradigms (Harter et al., 1984; McGinnis et al., 2011) and monitoring paradigms (Hillyard et al., 1998; Keil and Müller, 2010). In rare target detection paradigms, which are used in the current study, the subject must respond to an infrequent target stimulus defined by a conjunction of features (e.g. red square). Non-target stimuli that share a feature value with the target (e.g. red circle or blue square) elicit a larger SN than non-targets with no target features (e.g. blue circle). Because the SN is elicited by stimuli with attended vs. unattended features, has a scalp distribution that varies depending on modality and visual sub-modality (Anllo-Vento and Hillyard, 1996; McGinnis and Keil, 2011), has been localized to occipito-temporal visual cortex (Schoenfeld et al., 2007), it is believed to index attentional modulation of perceptual cortex (Hillyard and Anllo-Vento, 1998; McGinnis and Keil, 2011).

Another ERP component that is sensitive to targets is the N250, which has a latency and scalp distribution quite similar to the later portion of the selection negativity (Tanaka et al., 2006). In contrast to the SN, the N250 is typically elicited by complex objects, like faces (Tanaka et al., 2006) or animals (Pierce et al., 2011). Interestingly, the N250 is also elicited by rare target stimuli or stimuli that were previously targets (Tanaka et al., 2006), potentially linking the component to the SN. The N250 has also been observed to be larger in stimuli associated with subordinate level categorization training, even when those stimuli are not targets (Scott et al., 2006, 2008). To our knowledge, no study of the N250 has manipulated similarity between target