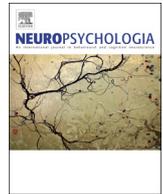




ELSEVIER

Contents lists available at ScienceDirect

Neuropsychologia

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

Color and spatial frequency differentially impact early stages of perceptual expertise training

Hélène Devillez^{a,*}, Matthew V. Mollison^a, Simen Hagen^b, James W. Tanaka^c, Lisa S. Scott^d,
Tim Curran^a

^a University of Colorado Boulder, United States

^b Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, France

^c University of Victoria, Canada

^d University of Florida, United States

ARTICLE INFO

Keywords:

Object categorization
Event-related potentials
Perceptual expertise
Perceptual learning
Spatial frequency
Color

ABSTRACT

The current study examined the role of color and spatial frequency on the early acquisition of perceptual expertise after one week of laboratory training with bird stimuli. Participants learned to categorize finches (or warblers) at the subordinate species level (e.g., purple finch) and categorize warblers (or finches) at the more general family level. Training images were presented in their natural colors across 6 sessions. Participants completed a subordinate level species matching task prior to training, one day after training and one week after training while event-related potentials (ERPs) were recorded. Bird images were presented in either their natural congruent color, incongruent color, grayscale, low spatial frequency (LSF < 8 cycles per image) or high spatial frequency (HSF > 8 cycles per image). Replicating previous training studies, performance benefited more from subordinate- than basic-level training. Before training, any color helped performance, but color congruence effects (congruent > incongruent) only emerged after subordinate-level training. Spatial frequency manipulations did not interact with training. The N170 ERP component was sensitive to spatial frequency manipulations, but not color. N170 spatial frequency effects did not interact with training, and training effects generalized to all manipulations except the LSF images. Like performance, color congruence effects on the N250 were only observed after subordinate level training. These results are consistent with previous reports suggesting that effects of perceptual expertise training on performance are more clearly indexed by N250 than N170 effects. Taken together, our behavioral and ERP results show that color plays an important role in both low- and high- level visual processing, supporting surface-plus-edge-based theories for object processing and recognition.

1. Introduction

Although any object can be identified at multiple levels of abstraction, it has been shown that objects are typically recognized fastest at the basic level (e.g., bird, car, chair) (Rosch et al., 1976). In contrast, object recognition is slower at levels that are either more superordinate (e.g., animal, vehicle, furniture) or more subordinate (e.g., sparrow, Honda, rocking chair) to the basic level (Rosch et al., 1976; Mack et al., 2009). However, with experience and practice, experts (e.g., bird-watchers or car experts) demonstrate a “downward shift” in their recognition where objects are recognized as quickly at the subordinate level as the basic level (Tanaka and Taylor, 1991; Johnson and Mervis, 1997). Bird experts, for example, are as fast to recognize a bird as a “sparrow” as they are to recognize the same image as a “bird”. In the

present study, we investigated the role of color and spatial frequency (SF) on the early acquisition of perceptual expertise after one week of laboratory training with bird stimuli.

Real-world expertise has been investigated across a variety of domains including car (Gauthier et al., 2000; Grill-Spector et al., 2004; Rossion et al., 2007; Stein et al., 2016) and bird experts (Johnson and Mervis, 1997; Gauthier et al., 2000; Tanaka and Curran, 2001; Martens et al., 2018). Real-world experts acquire their expertise on the scale of years, but it is possible to approximate the early acquisition of expertise by training participants in the laboratory across several days with novel objects (Gauthier et al., 1999; Cheung and Gauthier, 2014; Jones et al., 2018), or real-world objects, like cars (Scott et al., 2008) or birds (Tanaka et al., 2005; Scott et al., 2006). In previous expertise training studies, participants were trained to classify objects at either the basic

* Correspondence to: Department of Psychology and Neuroscience, Muenzinger Psychology Building, 345 UCB, Boulder, CO 80309-034, United States.

E-mail address: helene.devillez@colorado.edu (H. Devillez).

(e.g. wading birds versus owls) or the subordinate levels (e.g., Snowy Egret versus White Ibis). Only the participants who received practice with subordinate-level identification recognized these birds equally fast at the subordinate and basic levels and their speeded subordinate level recognition transferred to new instances from within the expert category (Tanaka et al., 2005).

Event-related potentials (ERPs) provide a reliable neural marker of real-world (Tanaka and Curran, 2001; Busey and Vanderkolk, 2005) and laboratory-trained (Scott et al., 2006, 2008) perceptual expertise. Training with visual objects enhances two distinct occipitotemporal ERP components, the N170 and N250 which peak around 170 and 250 ms respectively (Tanaka and Curran, 2001; Scott et al., 2006, 2008). The N170 component is thought to index some aspects of face processing (Bentin et al., 1996). Carmel and Bentin (2002) found a larger N170 component elicited for faces compared to cars, birds or items of furniture. Tanaka and Curran (2001) showed that the N170 component was also enhanced when bird or dog experts categorized objects within their domain of expertise (e.g., birds) compared to another domain (e.g., dogs). This result has been replicated with car experts (Gauthier et al., 2003). In addition, an N170 inversion effect, similar to what is seen with faces (for review see Rossion et al., 2002), was reported for fingerprint recognition in experts such that the N170 component was delayed for inverted fingerprints in experts but not in novices (Busey and Vanderkolk, 2005). The N170 component has also been shown to be modulated during tasks involving real-world objects or computer-generated artificial stimuli with experimentally induced expertise. In two different studies, adults learned to classify wading birds and owls (Scott et al., 2006) or three different categories of cars (Scott et al., 2008). They reported that both basic- and subordinate-level training enhanced the N170 component. Curran et al. (2002) similarly reported that the N170 component was enhanced after training to recognize categories of visually similar novel objects called “blobs”. An inversion effect, similar to what is seen with faces and fingerprint experts, was reported after training participants to recognize computer-generated artificial stimuli called “Greebles” (Rossion et al., 2002). Taken together, these results suggest that the N170 component is not necessarily face-specific, but modulated by perceptual experience and expertise.

The N250 component has also been reported to be associated with face processing (Schweinberger et al., 2002), to be sensitive to repetition of familiar celebrity identities (Schweinberger et al., 2004), one's own face and a recently familiarized face (Tanaka et al., 2006) as well as to be enhanced after learning for non-repeated novel faces (Kaufmann et al., 2009; Limbach et al., 2018). In object recognition, the N250 is elicited by personally familiar objects (i.e., own car, own dog) (Pierce et al., 2011). Whereas the N170 component is primarily sensitive to object categories, the N250 is sensitive to the familiarity of exemplars within the face or object categories. Scott et al. (2006, 2008) found that only subordinate-level training enhanced the N250 component. Subordinate-level training of other-race faces (individual level) also enhanced the N250 component compared to basic-level training (race level) (Tanaka and Pierce, 2009). Folstein et al. (2017) had participants practice a categorization task that included artificial animal stimuli followed by a speeded target detection task with both trained and untrained stimuli. They reported that trained stimuli elicited a significantly larger N250 than untrained stimuli. Overall, these results suggested that the N250 component is sensitive to individuation and repetition of unfamiliar faces but also generalizes across different representations of the same object and reflects subordinate-level access to object or face representations.

Both color and shape information play a critical role in object recognition (Biederman, 1987; Gauthier and Tarr, 1997; Tanaka and Presnell, 1999; Collin and McMullen, 2005). Color improves object recognition (Wurm et al., 1993) and contributes in a small way to object recognition with long-term color knowledge playing an important top-down role (Mapelli and Behrmann, 1997). It benefits for delineating

shape and surface segmentation in early visual processing (Gegenfurtner and Rieger, 2000). Gegenfurtner and Rieger (2000) showed that natural scenes were better encoded when presented with color relative to grayscale and suggested that color plays an important role at the very earliest stages of analysis of a scene to recognize and memorize it. In the same vein, objects associated with a particular color (high-color diagnosticity, e.g., banana) are recognized faster when they are presented with their natural congruent color than when they are presented in grayscale or with incongruent colors, while objects not associated with a particular color (low-color diagnosticity, e.g., car) are recognized equally fast when presented with color, in grayscale, or with incongruent colors (Tanaka and Presnell, 1999; Nagai and Yokosawa, 2003). It has been demonstrated that extensive real-world experience with a color diagnostic object domain can influence the role of color cues in subordinate recognition (Hagen et al., 2014). Here, expert and novice birdwatchers were tested using colored, grayscale, and incongruent bird images while having to recognize the birds at the subordinate levels (e.g., “Robin”). Although color congruence affected both the fast and the slow recognition judgments of the experts, it affected only the slow judgements of the novices, suggesting that while the use of color cues were automatic in the experts, it was more deliberate in the novices. Thus, real-world experience acquired by the bird experts were associated with the ability to instantly rely on diagnostic color cues.

The role of shape information in recognition depends on the category level at which the objects are recognized. Different types of shape information can be isolated by filtering images for different ranges of SFs. For example, the low resolution of low SFs (LSFs) carries coarse shape information (e.g., global shape and internal blobs) and the high resolution of high SFs (HSFs) carries fine-grained information (e.g., fine edged contours, internal edge shapes). Subordinate-level recognition is reported to depend disproportionately on HSF information whereas basic-level recognition relies more on LSF information (Collin and McMullen, 2005), suggesting that the former category-level relies more on fine-grained features such as internal feature information. Similar to color, it has been demonstrated that extensive real-world object experience can influence the role of internal shape information in subordinate recognition (Hagen et al., 2016). Specifically, expert and novice birdwatchers were tested with bird images that were band-pass filtered over a range of different SFs (low to high SFs), while masking the external contour so to only vary internal shape information. While both the experts and the novices were more accurate when bird images contained the internal shape information carried by a midrange of SFs (8–32 cycles per image - cpi), only the experts were faster in the same midrange. Thus, the real-world experience did not change the range of shape information that was used, but instead influenced the efficiency with which the observers could utilize shaped in the midrange. However, it is unknown whether these changes develop early or late in the acquisition of real-world expertise.

Whereas the study of extant, real-world experts sheds light on important perceptual cues, such as color and SFs, in expert recognition, laboratory training studies show how these cues are utilized during the acquisition of expertise. In a recent investigation, Jones et al. (2018) examined the impact of color and SF manipulations (similar to Hagen et al., 2014, 2016) before and after adults were trained to categorize an artificial creature species. Adults were trained to categorize one object family at the basic level and another family at the subordinate level. Subordinate-level matching of both families was tested before and after training with images that appeared in colors that were congruent with training, grayscale images, or images retaining either low SFs (LSF < 8 cpi) or higher SFs (HSF > 8 cpi). As in previous training experiments, discrimination increased after subordinate- but not after basic-level training. These training effects generalized to each of the three image manipulations. Jones et al. (2018) also measured ERPs before and after training. Both the N170 and N250 amplitudes were enhanced more after subordinate- than basic-level training, departing from previous

research showing that training level did not influence N170 training effects (Scott et al., 2006, 2008). Both components were also influenced by the image manipulations, but these manipulations did not interact with training. As discussed by Jones et al. (2018), the use of novel objects might account for differences from previous experiments using categories that adults have previous experience with, including birds (Scott et al., 2006) and cars (Scott et al., 2008).

In the present study, adults were trained to categorize ten species of birds from the finch family and ten species of birds from the warbler family. Similar to Hagen et al. (2014, 2016), we manipulated color and SF to examine the effects on perceptual expertise training with birds. Following previous studies (Scott et al., 2006, 2008; Jones et al., 2018), we defined basic level at the family level of warbler and finch and subordinate level at the species level (e.g., Bachman warbler, Bay-crested warbler). Participants were trained to categorize the finches (or warblers) at the subordinate species level (e.g., purple finch) and categorize the warblers (or finches) at the more general basic level. Training images were presented in their natural colors across 6 sessions. Participants were given a subordinate level species matching task prior to training, one day after training and one week after training. In this task, participants decided whether two sequentially presented bird images belonged to same or different species. The bird images were presented in either their natural color (congruent condition), incongruent color, grayscale, low SF (LSF < 8 cpi) or high SF (HSF > 8 cpi). We also tested both the originally trained image exemplars as well as untrained exemplars of the trained species to examine generalization beyond the training set.

Based on past expertise studies, we made several predictions about how color and SF information might impact recognition after basic-versus subordinate-level training. In general, because color and SF effects seen in real-world experts emerged after some experience, we expected subordinate training to yield color and SF effects that are consistent with those observed in bird experts (Hagen et al., 2014, 2016). Because color plays a role in early visual processes of edge detection and texture segmentation (Cavanagh, 1987; Gegenfurtner and Rieger, 2000), we predicted that both congruent and incongruent color would produce better performance than grayscale images before training. However, because congruent, but not incongruent, color improves performances of species categorization for real-world bird experts (Hagen et al., 2014), we expected that only subordinate level training produces the color-congruency effect (congruent > incongruent). With respect to SF manipulations, we hypothesized that overall, the full spectrum images would produce the best recognition followed by the HSF images, which preserve more useful detail information than LSF images (Collin and McMullen, 2005; Hagen et al., 2016; Jones et al., 2018). We expected that training may increase behavioral performance advantage for HSF over LSF. Hagen et al. (2016) reported that experts showed larger SF effects on reaction time than novices but did not report accuracy difference between experts and novices. However, in their study, they used an easier task than in the present study so we hypothesized that training will improve accuracy in a more difficult task.

We expected to replicate the N170 and N250 ERP results from previous training experiments with birds (Scott et al., 2006) and cars (Scott et al., 2008). In particular, these studies indicated that the N250 effects parallel behavioral findings more so than the N170. Just as performance improvements follow subordinate-but not basic-level training, the N250 amplitude should only be enhanced by subordinate training. On the other hand, as a marker of category exposure, the N170 should be higher amplitude at post- than pre-test, but equally enhanced by subordinate- and basic-level training (Scott et al., 2006, 2008). Furthermore, we expected any image manipulation effects on performance (e.g., the predicted dependence of color congruency effects on subordinate level training) to more likely be reflected in N250 than N170 amplitudes, based on previous indications that the N250 is more consistently related to behavior.

2. Method

2.1. Participants

Seventeen novice participants, ranging from 18 to 29 years of age ($M = 22.17$, $SD = 0.79$, 9 female, 1 Hispano or Latino, 7 Asian, 9 White) were trained to recognize several exemplars of several different bird species. Sample size was selected to be consistent with previous research (Scott et al., 2006, $n = 16$; Scott et al., 2008, $n = 12$). To compensate for participant attrition, 26 adult participants were recruited in total. Within this group, the data from three additional participants were lost due to technical issues. One additional participant was stopped early due to poor performance. Five additional participants dropped out before completing all the training sessions. Participants had no prior experience in bird-watching before the training.

Participants received monetary compensation for their participation. All participants were right-handed native-English speakers and had normal or corrected-to-normal vision. Informed consent was obtained from each participant, and the study was approved by the University of Colorado Boulder Institutional Review Board (IRB).

2.2. Stimuli

Stimuli included 12 images for each of the 10 species from the finch family (Pine Grosbeak, Pine Siskin, American Gold, House, Evening Grosbeak, Common Redpoll, Cassins, Brambling, Black Rosy, White Winged Crossbill) and 12 images for each of the 10 species from the warbler family (Nashville, Golden Winged, Chestnut Sided, Canada, Blue Winged, Black Throated Green, Blackburnian, Bay Breasted, American Redstart, Wilsons) (Fig. 1).

Using Matlab,¹ the images were transformed to manipulate the color and SF content. Birds were manipulated during the matching task to appear in one of five conditions: color congruent, incongruent, grayscale – all SF (AllSF), low SF (LSF < 8 cpi) and high SF (HSF > 8 cpi). The bird images were transformed to create an incongruent condition and a grayscale condition, using the L^*a^*b color space. This color space has been used in previous studies investigating the role of color scene and expert bird recognition (Oliva and Schyns, 1997; Hagen et al., 2014). The incongruent condition was created by either rotating the color axis (e.g., red to green or blue to yellow or vice versa), by swapping the two color axes (e.g., blue to red), or by both swapping and flipping the color axes. The decision of which transformation to use depended on which transformation created the subjectively best incongruent condition. The grayscale bird images were created by eliminating the hue and saturation information while retaining the luminance. For the SF manipulations, the grayscale bird images were normalized to the mean then band-pass filtered for two ranges of SFs: > 8 cpi and < 8 cpi. We chose a cutoff of 8 cpi based on previous studies on object recognition (Collin and McMullen, 2005; Harel and Bentin, 2009; Hagen et al., 2016). A mask of the external bird contour was applied to keep the external contour constant for the two SF ranges (Hagen et al., 2016). Fig. 1 illustrates the stimuli and the transformations used in this experiment. Images were cropped and scaled to fit within a frame of 450×450 pixels and pasted on a gray background. Images subtended a visual angle of approximately 8.79 vertically and 8.90 horizontally.

2.3. Procedure

Participants performed a pretest matching task, then 6 training sessions were completed and finally participants performed a posttest matching task one day after the last training session (posttest) and one week after the last training (delay).

¹ <https://github.com/warmlogic/expertTrain/tree/master/imageProcessing>.

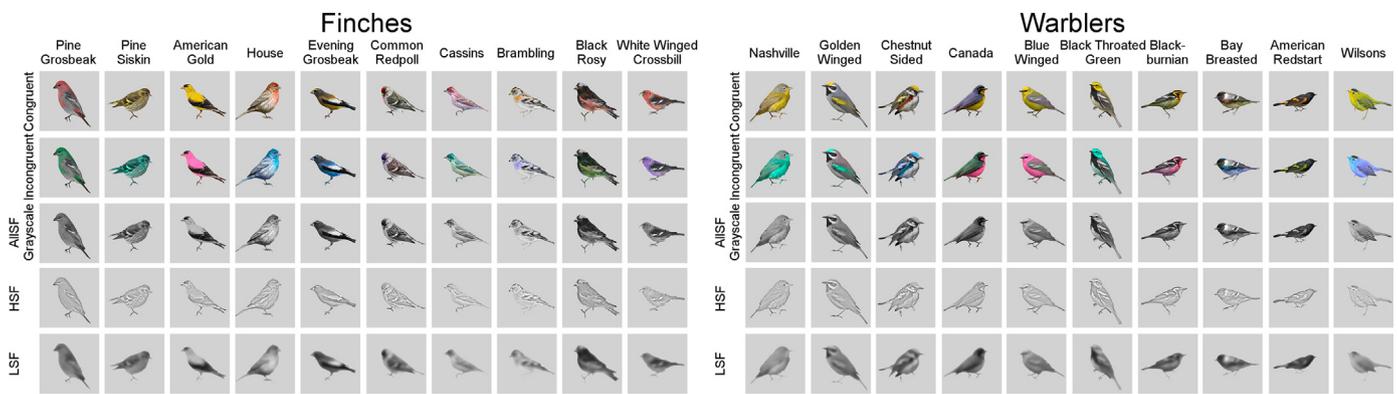


Fig. 1. Examples of the stimuli used, for the two families Finches and Warblers and the ten species for each family. The two top rows show the color manipulation: congruently colored birds and incongruently colored birds. Middle row shows the grayscale versions. The two bottom rows show the SF manipulation: HSF (> 8 cycles per image) and LSF (< 8 cycles per image).

2.3.1. Matching task

Participants completed a subordinate-level sequential matching task that has previously been shown to be sensitive to differences in levels of perceptual expertise (Gauthier et al., 2000; Scott et al., 2008). Participants were shown a fixation point for 500–700 ms then a stimulus for 800 ms followed by a fixation point for 1000–1200 ms and then another image for an additional 800 ms (Fig. 2). Then the participants were immediately presented with a question mark and were required to indicate whether the two images were of the SAME (e.g. two “Nashville Warblers”) or of a DIFFERENT (e.g., a “Nashville Warbler” and a “Wilson’s Warbler”) species. SAME trials were always different exemplars of the same species. DIFFERENT trials included two exemplars of different species within the same family. The question mark remained on the screen until a response was made. All stimuli were randomly ordered and randomly matched within each condition. This task consisted of 1200 trials.

Across both same and different trials, there were four different types of trials: 1) Trained Exemplar/Trained Exemplar (350 trials), 2) Untrained Exemplar/Untrained Exemplar (350 trials). Two other conditions were presented by mistake: Trained Exemplar/Untrained Exemplar (250 trials) and Untrained Exemplar/Trained Exemplar (250 trials). The latter two mistaken trial types were excluded from analyses because they were not part of the intended design. Note that these two conditions were presented to all participants, which means that if ever there is an influence on the results, it would be the same for all participants and all conditions. The trained exemplars were specific exemplars from each of the bird families used during training. The untrained exemplars included new instances of each species of the two bird families. Assignment of exemplars to test conditions was counter-balanced across subjects.

2.3.2. Training task

Half of the participants learned to differentiate birds from the finch (or warbler) family at the subordinate level of species (e.g., “purple finch”), whereas birds from warbler (or finch) family was categorized at the basic level of family (e.g., “warbler”). The exact same stimulus pairs were shown in the pre- and post-tests, but in different randomized orders.

Participants were shown a fixation cross for 500–700 ms then a stimulus for 1000 ms followed by a question mark for 2000 ms. They were then required to identify the species the bird was from (species 1–10) or whether the bird was from the other family (Fig. 2). Species numbers 1–10 were randomly assigned to the subordinate-level family, for each participant. Species were labeled with arbitrary numbers rather than actual species names to facilitate manual responding and to minimize subject differences in familiarity with the species names. The number labeling technique was not meant to teach the participants the proper names for the birds, but to associate each bird species with a unique key press. This type of labeling technique has been successfully applied in other training studies (e.g., Tanaka and Curran, 2001; Scott et al., 2008). Feedback was given on every trial by an auditory tone indicating a correct or incorrect response and the bird’s species number was shown either in green (correct responses) or red (incorrect response). For each family, 6 exemplars for each of the 10 species were trained. Birds were shown in their normal colors during training. All stimuli were randomly ordered, with the constraint that no more than 3 exemplars from the same family could occur in a row.

The first training session differed from the subsequent 5 sessions in that it consisted in 240 trials preceded by a phase that slowly introduced species from both families, rotating through in a total of 5 blocks. The other training session consisted of 480 trials, divided in 4 phases consisted of naming each exemplar once (120 trials), so a

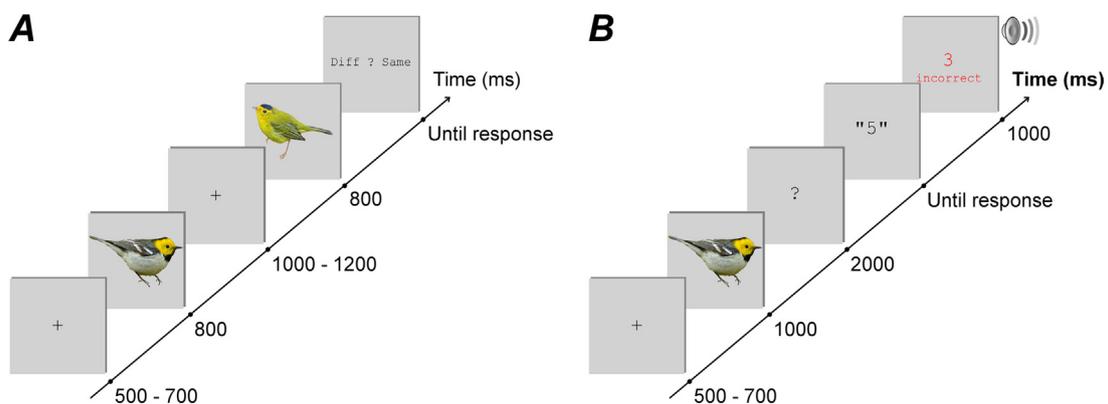


Fig. 2. Time course of one trial during the matching test (A) and during the training task (B).

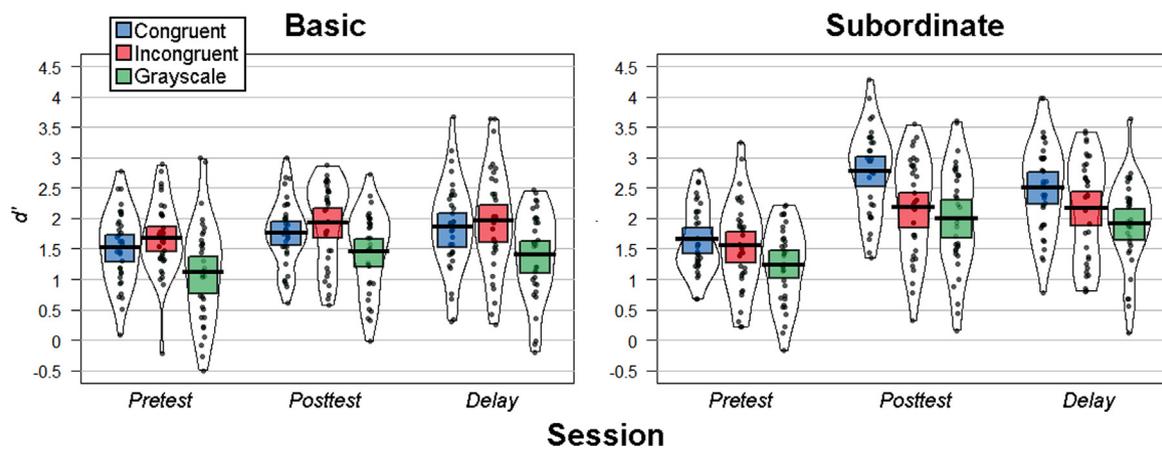


Fig. 3. Matching performance d' for pretest, posttest and delay, for congruent, incongruent and grayscale images, collapsed over trained and untrained exemplars. Colored regions represent the 95% confidence interval. Thick black lines represent the mean. Individual data points represent the mean d' for each participant.

participant had the chance to name each exemplar 4 times in a session (480 trials).

2.3.3. EEG recordings

A 128-channel HydroCel Geodesic Sensor Net TM (GSN 200, v. 2.1; Tucker, 1993) was used to measure the EEG at the scalp using a central vertex reference (Cz) with a sampling rate of 250 Hz, a high-pass hardware filter of 0.01 Hz, and a low-pass hardware filter of 100 Hz. The net was connected to an AC-coupled, high-input impedance amplifier (Net Amps TM; Electrical Geodesics, Inc., Eugene, OR) and recordings were made using the Net Station application. The electrodes were adjusted until impedance measurements were less than 40 k Ω . All data processing steps and analyses were done in MATLAB using in-house scripts calling functions from the FieldTrip toolbox (Oostenveld et al., 2011). A high-pass filter at 0.1 Hz, low-pass filter at 100 Hz, and a notch filter from 59 to 61 Hz were applied to the data. Study and test trials were epoched into 1200 ms segments, 200 ms before the onset of each stimulus and 1000 ms after. Artifact detection was used to reject particularly noisy epochs, as well as those that exceed an amplitude of $\pm 100 \mu\text{V}$. An average of 2.07% ($SD = 1.46$) of the trials were rejected. The data were referenced to the average of all channels and individual trials were baseline corrected relative to -200 – 0 ms. EEG was recorded during both the training and testing phase, but given the goals of the present investigations analyses reported here focused on the three test phases (pretest, posttest and delay).

2.4. Analysis

The effects of color and SF manipulations were analyzed separately. Note that the “grayscale” label in the color analyses refers to exactly the same trials as the “AllSF” label in the SF analyses, to emphasize the most pertinent attribute of those trials in each set of analyses.

d' and RT were separately analyzed in a mixed-design analysis of variance (ANOVA) with manipulation – color or SF (congruent, incongruent, grayscale or AllSF, HSF, LSF), session (pretest, posttest, delay), level (basic, subordinate) and exemplar type (trained, untrained) as within-subject factors. Multiple comparisons were assessed with Fisher LSD post-hoc tests.

Based on previous perceptual expertise studies (Scott et al., 2006, 2008; Jones et al., 2018), ERP analyses focused on the visual P1, N170 and N250 components. Exploratory P1 analyses were conducted after observing apparent P1 effects in our grand averages. For each component, we averaged over electrode montages at left occipito-temporal (50 51 57 58 59 64 65) and right occipito-temporal (90 91 95 96 97 100 101) regions. These clusters were selected based on previous studies (Scott et al., 2006, 2008). P1 amplitude was calculated as the mean

amplitude over the time window spanning 97–141 ms using peak detection (the window was centered on the peak of the grand average ± 2 standard deviations), N170 over the time window spanning 155–211 ms (Scott et al., 2006), and N250 over the time window spanning 230–330 ms (Scott et al., 2006, 2008). The data of one participant was not used for the ERP analyses due to low trial counts. P1, N170 and N250 amplitude were separately analyzed in a mixed-design ANOVA with manipulation – color or SF (congruent, incongruent, grayscale or AllSF, HSF, LSF), session (pretest, posttest, delay), level (basic, subordinate), exemplar type (trained, untrained), hemisphere (left, right) and stimulus presentation order (or “order”) (stim1, stim2) as within-subject factors. Although we have no hypotheses concerning order, it is included to account for additional variability and for consistency with previous work (Scott et al., 2006, 2008). Hence, order effects are reported in ANOVA tables without discussion. Multiple comparisons were assessed with Fisher's LSD post-hoc tests. An average of 54.64 ($SD = 0.87$, min = 23) trials (out of 60) per condition were included for each condition across participants.

Because of *a priori* interest in whether or not training effects would be observed when each of the image manipulations was considered alone, we ran separate ANOVAs for each color and SF manipulation, for d' , P1, N170 and N250 with session (pretest, posttest, delay) and level (basic, subordinate) as within-subject factors.

Details of main effect and interaction statistics from the ANOVA analysis are reported in tables, in the [Supplementary material section](#). p -values are reported in the text, denoted by p for main effects and interactions from the ANOVA and p_{Fisher} for Fisher's LSD post-hoc tests.

3. Results

3.1. Color manipulation

3.1.1. Behavioral performance

3.1.1.1. Accuracy (d'). Fig. 3 presents the d' matching performance for pretest, posttest and delay for the congruent, incongruent and grayscale conditions. Complete results of the statistical analysis are presented in [Table A. 1](#). Analysis of accuracy indicated a significant main effect of session, $F(2,32) = 18.20$, $p < .001$, showing that participant's accuracy increased with training (pretest < posttest and pretest < delay, both $p_{\text{Fisher}} < .001$). The main effect of condition, $F(2,32) = 23.69$, $p < .001$, revealed that participants were more accurate for congruent and incongruent color compared to grayscale images (both $p_{\text{Fisher}} < .001$). We also observed a main effect of level, $F(1,16) = 15.66$, $p < .01$, with better accuracy for the subordinate compared to the basic level. The significant condition \times level interaction, $F(2,32) = 14.63$, $p < .001$, showed that the d' difference

between the subordinate and basic levels was greater for congruent ($p_{Fisher} < .001$) and grayscale ($p_{Fisher} < .01$) images compared to incongruent images. The significant level \times exemplar interaction, $F(1,16) = 5.51, p < .05$, indicated that, at the subordinate level, participants were more accurate for trained stimuli compared to untrained stimuli ($p_{Fisher} < .05$), but no difference was observed at the basic level. The significant session \times level interaction, $F(2,32) = 12.64, p < .001$, revealed that participants' session differences (posttest $>$ pretest and delay $>$ pretest) were larger following subordinate than basic training (both $p_{Fisher} < .01$). When each of the color conditions was analyzed in separate ANOVAs, session effects were significant for each color condition (Congruent: $F(2,32) = 23.18, p < .001$, Incongruent: $F(2,32) = 6.63, p < .01$ and Gray: $F(2,32) = 13.81, p < .001$).

Following the significant session \times level interaction, we ran separate subordinate and basic level ANOVAs (complete results are presented in Table A. 2). At the basic level, the significant main effect of session, $F(2,32) = 5.70, p < .01$, showed that participants were more

accurate after training (pretest $<$ posttest and pretest $<$ delay, both $p_{Fisher} < .01$). The main effect of condition, $F(2,32) = 15.89, p < .001$, revealed that participants were more accurate for congruent and incongruent compared to grayscale images (both $p_{Fisher} < .001$). At the subordinate level, analysis of accuracy indicated a main effect of color, $F(2,32) = 30.29, p < .001$, with grayscale images showing lower accuracy than congruent and incongruent images (all $p_{Fisher} < .001$). This was qualified by a significant color \times session interaction, $F(4,64) = 2.72, p < .05$, such that accuracy for congruent and incongruent images did not differ before training, but congruent images showed a better accuracy than incongruent images after training ($p_{Fisher} < .001$). Furthermore, subordinate-level training effects were significant after both the posttest and delay session, for each of the three color conditions (all $p_{Fisher} < .01$).

3.1.1.2. Reaction time. A plot of response time is presented in Fig. A. 1 and complete results of the statistical analysis are presented in Table A. 3. The main effect of condition, $F(2,32) = 7.70, p < .01$, revealed that

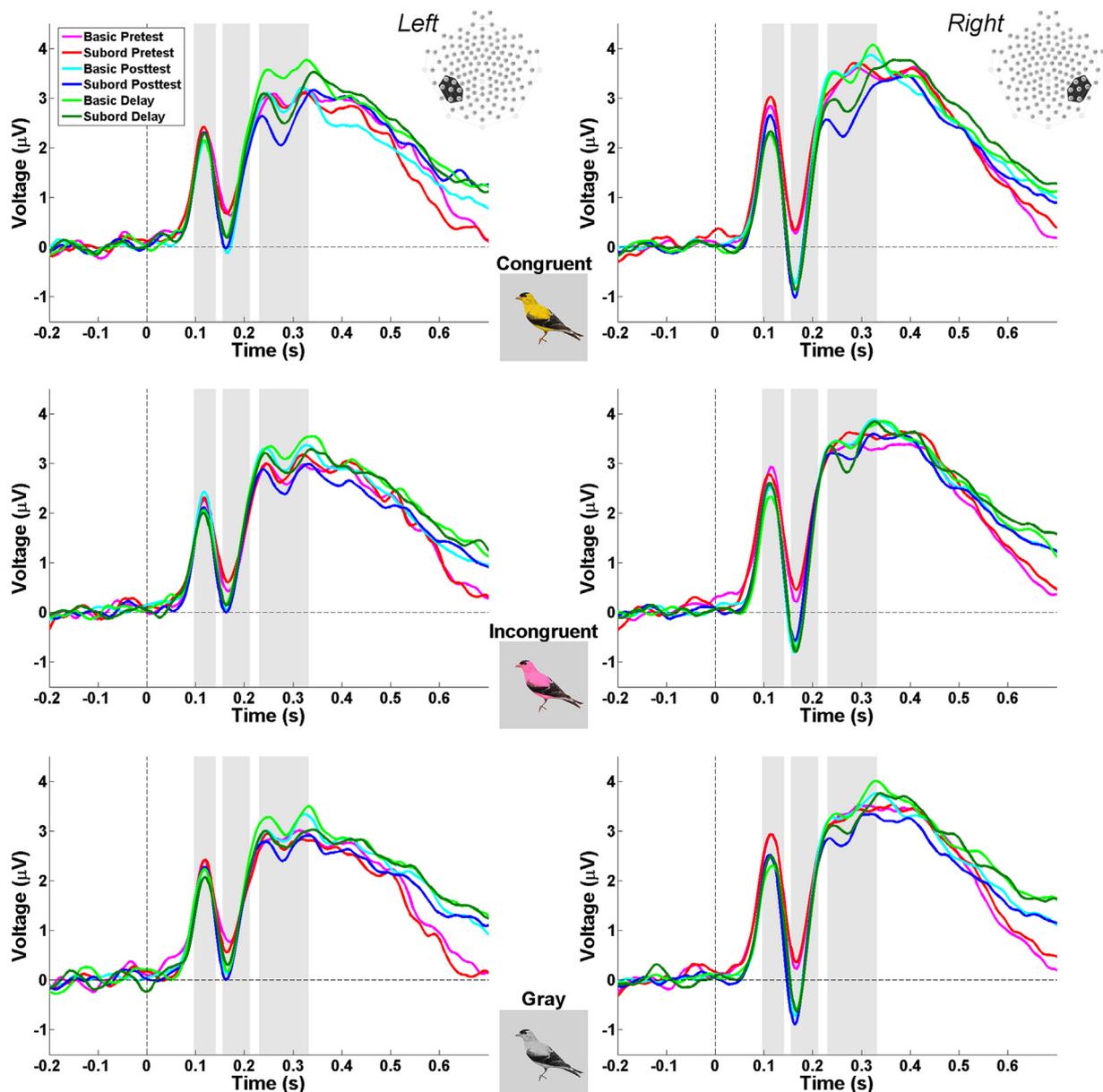


Fig. 4. ERP for pretest, posttest and delay, for congruent, incongruent and grayscale images, collapsed over exemplar types and order. Gray regions show time windows for the P1, N170, and N250 ERP components.

participants were slower for grayscale images compared to both congruent and incongruent images (both $p_{Fisher} < .05$).

3.1.2. ERP results

Fig. 4 presents the ERP data for pretest, posttest and delay, for the three color conditions.

3.1.2.1. P1 component analysis. A plot of the amplitude of the P1 ERP is presented in Fig. A. 2 and complete results of the statistical analysis are presented in Table A. 4. Analysis of P1 amplitude indicated a main effect of session, $F(2,30) = 3.85$, $p < .05$, showing that the amplitude of the P1 was larger during pretest compared to posttest and delay (both $p_{Fisher} < .05$). The significant color \times hemisphere \times session interaction, $F(4,60) = 3.30$, $p < .05$, indicated that the amplitude of the P1 was larger over the right than left hemisphere ($p_{Fisher} < .01$) and that session effects were greatest for congruent images over the right hemisphere ($p_{Fisher} < .01$). The color \times hemisphere \times session \times level interaction, $F(4,60) = 2.55$, $p < .05$, appeared to have captured slight differences in the pre-training, right hemisphere, color effect between subordinate and basic training levels. When each of the color conditions were analyzed in separate ANOVAs, the main effect of session was significant for congruent images, $F(2,30) = 4.95$, $p < .05$, with pretest $>$ posttest = delay (both $p_{Fisher} < .05$) and marginal for incongruent, $F(2,30) = 3.12$, $p = .059$, and gray, $F(2,30) = 3.17$, $p = .056$, images.

3.1.2.2. N170 component analysis. A plot of the N170 amplitude is presented in Fig. A. 3 and complete results of the statistical analysis are presented in Table A. 5. Analysis of N170 amplitude indicated a main effect of session, $F(2,30) = 4.45$, $p < .05$, qualified by a significant hemisphere \times session interaction, $F(2,30) = 3.98$, $p < .05$. Results showed that the right hemisphere N170 was smaller at pretest than posttest and delay (both $p_{Fisher} < .001$). For the left hemisphere, the N170 was smaller at pretest than posttest ($p_{Fisher} < .05$), but pretest and delay did not differ. As previously indicated, we have no hypotheses concerning order, so the last two significant interactions in Table A. 5 (order \times exemplar type and color \times session \times level \times exemplar type \times hemisphere \times order) are not discussed. When each of the color conditions were analyzed in separate ANOVAs, the main effect of session was significant for each color condition (Congruent: $F(2,30) = 4.82$, $p < .05$, Incongruent: $F(2,30) = 3.41$, $p < .05$ and Gray: $F(2,30) = 4.19$, $p < .05$), showing pretest $>$ posttest = delay (all $p_{Fisher} < .05$).

3.1.2.3. N250 component analysis. Fig. 5 shows the amplitude of the N250 ERP for pretest, posttest and delay, for the three color conditions. Complete results of the statistical analysis are presented in Table A. 6. Analysis of N250 amplitude indicated a main effect of level, $F(1,15) = 10.90$, $p < .01$, qualified by a significant color \times level interaction, $F(2,30) = 5.61$, $p < .01$, showing that the amplitude of the N250 was larger for subordinate- than basic-level for the three color conditions (all $p_{Fisher} < .001$). The session \times level, $F(4,60) = 2.90$, $p < .05$, and color \times session \times level interactions, $F(4,60) = 2.90$, $p < .05$, showed an increased N250 amplitude after training, for both posttest and delay, with subordinate-level training leading to a larger N250 than basic-level training. This is further discussed below. When each of the color conditions were analyzed in separate ANOVAs, the significant session \times level interaction showed that session effects were significant only at subordinate level for congruent, $F(2,30) = 12.30$, $p < .001$, with pretest $<$ posttest and pretest $<$ delay (both $p_{Fisher} < .01$) and incongruent, $F(2,30) = 7.88$, $p < .01$, with pretest $<$ posttest ($p_{Fisher} < .01$) images. For gray images, session effects were not significant.

Following the significant session \times level \times color interaction, we ran separate subordinate and basic level ANOVAs (complete results are presented in Table A. 7). At the basic level, the significant color

\times hemisphere \times exemplar type interaction, $F(2,30) = 3.49$, $p < .05$, showed that over the left hemisphere, the amplitude of the N250 was more negative for incongruent and gray untrained stimuli compared to congruent untrained images (both $p_{Fisher} < .05$) and that over the right hemisphere, the amplitude of the N250 was more negative for trained gray compared to trained congruent images ($p_{Fisher} < .05$). At subordinate level, the main effect of color, $F(2,30) = 3.38$, $p < .05$, revealed a larger amplitude of N250 for incongruent compared to congruent images ($p_{Fisher} < .05$). The significant color \times session interaction, $F(4,60) = 4.06$, $p < .01$, showed that the color effects differed before and after training. More specifically, as observed for d' , the N250 for congruent and incongruent images did not differ before training but amplitudes were more negative for congruent than incongruent images after training ($p_{Fisher} < .001$ for posttest session and $p_{Fisher} < .05$ for delay session). From a different perspective, N250 training effects were only observed for congruent images and did not generalize to incongruent and gray images. Furthermore, subordinate-level training effects were significant after both the posttest and the delay session in the congruent condition (both $p_{Fisher} < .001$), but only significant after posttest for incongruent and grayscale.

3.2. Spatial frequency manipulation

3.2.1. Behavioral performance

3.2.1.1. Accuracy (d'). Fig. 6 presents the matching performance data d' for pretest, posttest and delay for the three SF conditions. Complete results of the statistical analysis are presented in Table B. 1. As previously indicated, the grayscale results in the color analyses are based on exactly the same data as the AllSF results in the SF analyses. The main effect of session, $F(2,32) = 18.03$, $p < .001$, showed that participants' accuracy increased with training (pretest $<$ posttest = delay) (both $p_{Fisher} < .001$) and the main effect of SF, $F(2,32) = 64.58$, $p < .001$, revealed that all SF conditions significantly differed from each other (AllSF $>$ HSF $>$ LSF) (all $p_{Fisher} < .001$). We observed a main effect of level, $F(1,16) = 9.19$, $p < .01$, qualified by a significant SF \times level interaction, $F(2,32) = 3.35$, $p < .05$, indicating that the accuracy difference between the subordinate and basic levels was greater for AllSF compared to HSF and LSF images (both $p_{Fisher} < .001$). The significant level \times exemplar type interaction, $F(1,16) = 5.35$, $p < .05$, and the significant session \times level \times exemplar interaction, $F(2,32) = 3.31$, $p < .05$ revealed that no difference was shown between basic- and subordinate-level training for untrained stimuli but that participants were more accurate after subordinate training compared to basic training for trained stimuli ($p_{Fisher} < .001$). The significant session \times level interaction, $F(2,32) = 10.26$, $p < .001$, indicated that the training improvement was greater for subordinate than basic-level training ($p_{Fisher} < .001$ for both posttest and delay sessions). When each of the SF conditions were analyzed in separate ANOVAs, session effects were significant for each SF condition (AllSF: $F(2,32) = 13.81$, $p < .001$, HSF: $F(2,32) = 13.84$, $p < .01$ and LSF: $F(2,32) = 8.92$, $p < .01$).

Following the significant session \times level interaction, we ran separate subordinate and basic level ANOVAs (complete results are presented in Table B. 2). Within each ANOVA, the main effect of session, revealed that both the basic-, $F(2,32) = 3.29$, $p < .05$, and subordinate-, $F(2,32) = 34.47$, $p < .001$, level training showed the same training effect that was reported above in the initial ANOVA (pretest $<$ posttest = delay) (both $p_{Fisher} < .05$ for basic and both $p_{Fisher} < .001$ for subordinate). The main effect of SF for basic, $F(2,32) = 32.81$, $p < .001$, and subordinate, $F(2,32) = 41.38$, $p < .001$, also showed the same SF differences as reported previously (AllSF $>$ HSF $>$ LSF) (all $p_{Fisher} < .05$). Furthermore, at the subordinate level, the main effect of exemplar type, $F(1,16) = 4.90$, $p < .05$, showed that participants were more accurate with trained than untrained stimuli.

3.2.1.2. Reaction time. A plot of the response time is presented in Fig.

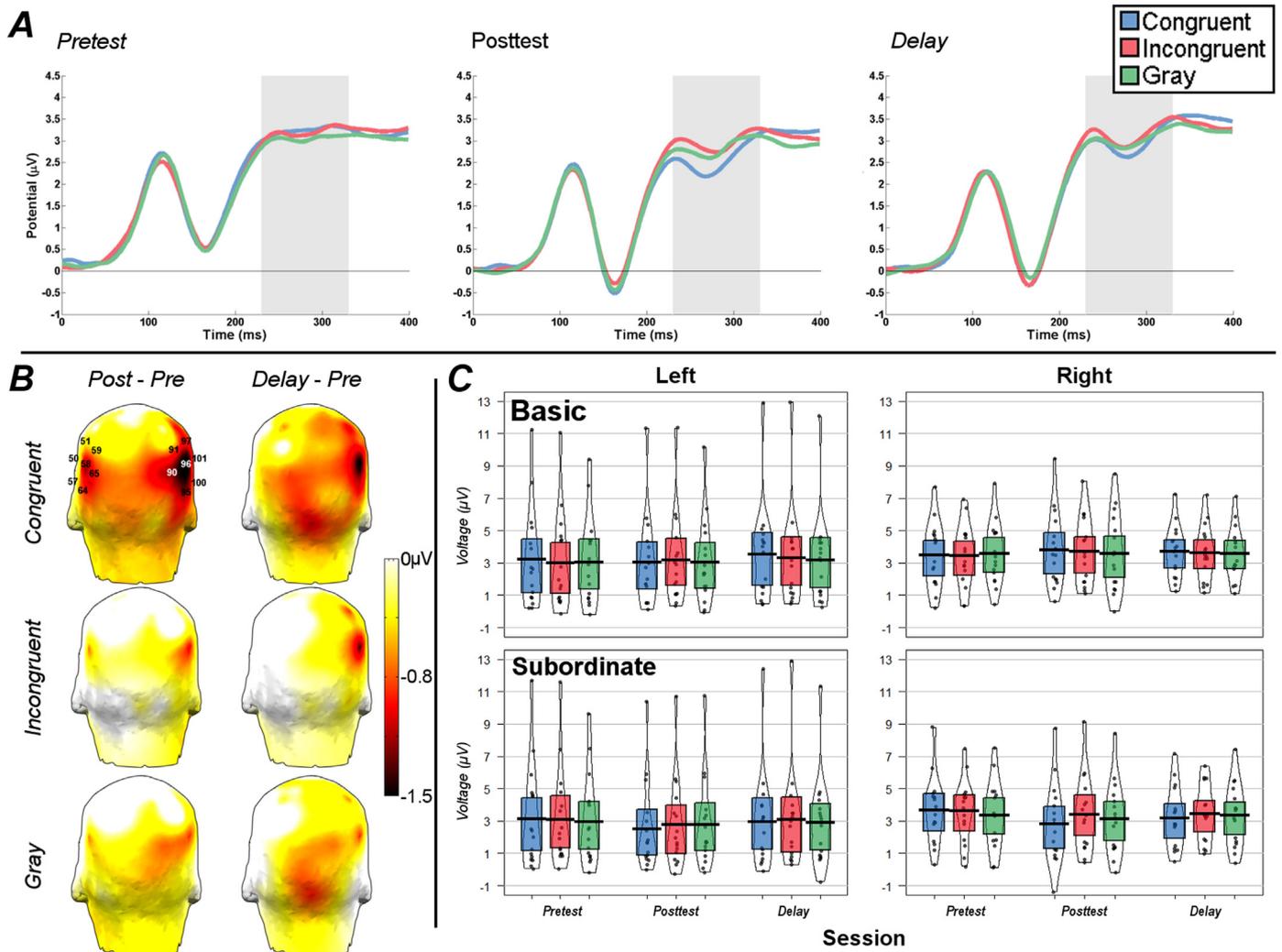


Fig. 5. A. N250 amplitude for pretest, posttest and delay, for congruent, incongruent and grayscale images and for subordinate training only, collapsed over hemispheres, order and exemplar types. Gray regions show time windows for the N250 ERP component B. Topography differences between posttest and pretest, and delay and pretest, for congruent, incongruent and grayscale images at time 280 ms C. N250 amplitude for pretest, posttest and delay, for congruent, incongruent and grayscale images, collapsed over order and exemplar types. Colored regions represent the 95% confidence interval. Thick black lines represent the mean. Individual data points represent the mean amplitude for each participant.

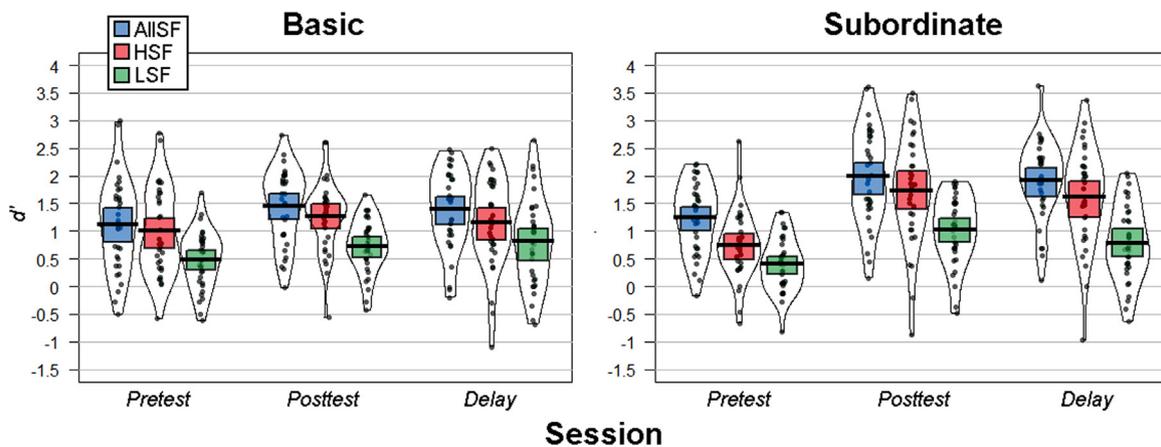


Fig. 6. Matching performance d' for pretest, posttest and delay, for AllSF, HSF and LSF images, collapsed over trained and untrained exemplars. Colored regions represent the 95% confidence interval. Thick black lines represent the mean. Individual data points represent the mean d' for each participant.

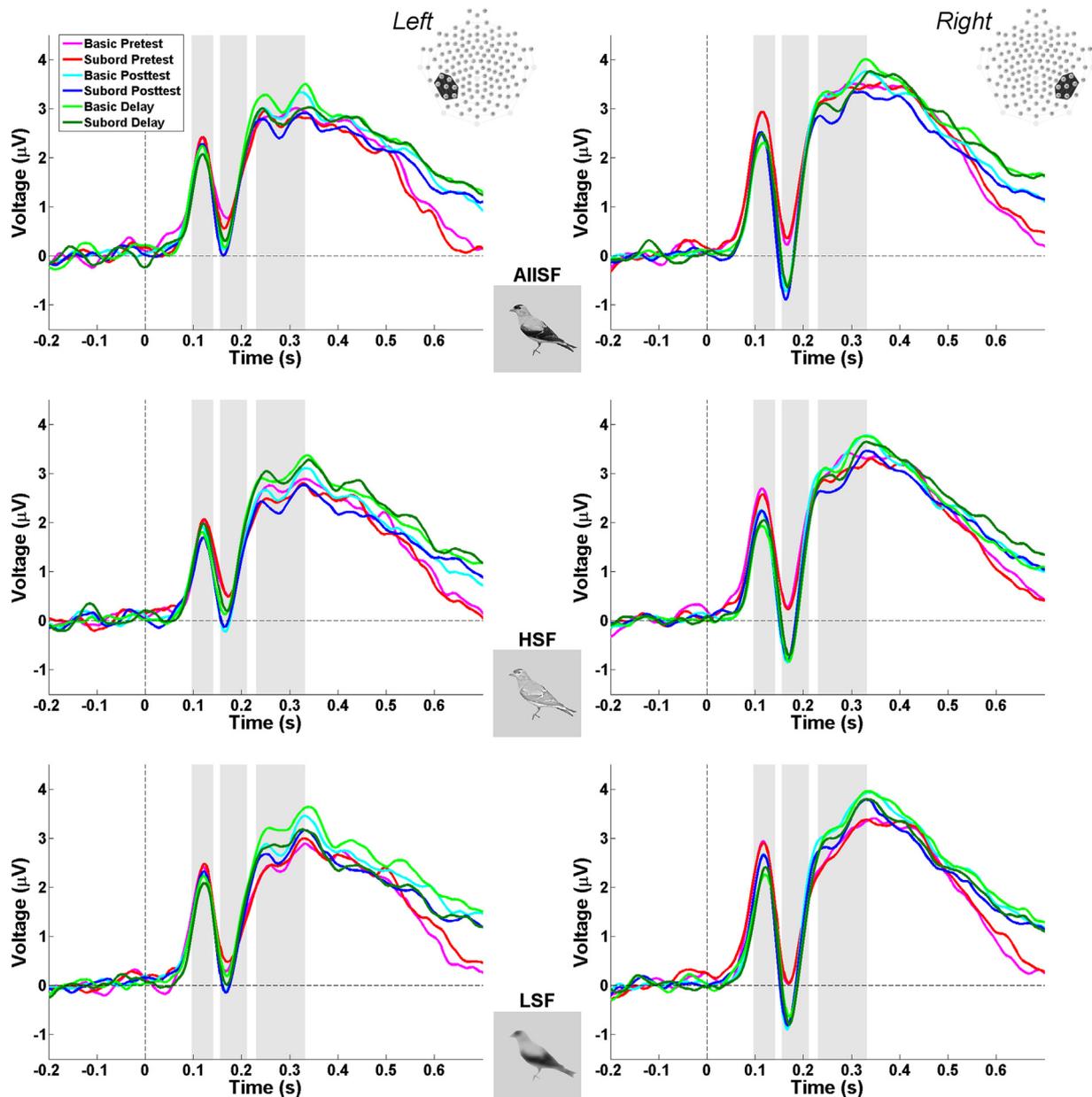


Fig. 7. ERP for pretest, posttest and delay, for AllSF, HSF and LSF images, collapsed over exemplar types and order. Gray regions show time windows for the P1, N170, and N250 ERP components.

B. 1 and complete results of the statistical analysis are presented in Table B. 3. The main effect of SF, $F(2,32) = 4.58$, $p < .05$, revealed that participants were slower for LSF images compared to AllSF images ($p_{Fisher} < .05$). The main effect of level, $F(1,16) = 6.91$, $p < .05$, showed that participants were faster for basic- compared to subordinate-level training.

3.2.2. ERP results

Fig. 7 presents the ERP data for pretest, posttest and delay, for the three SF conditions.

3.2.2.1. P1 component analysis. A plot of the amplitude of the P1 is presented in Fig. B. 2 and complete results of the statistical analysis are presented in Table B. 4. The main effect of SF, $F(2,30) = 17.31$, $p < .001$, revealed that the amplitude of the P1 was larger for AllSF and LSF images compared to HSF images (both $p_{Fisher} < .001$). The main effect of session, $F(2,30) = 3.72$, $p < .05$, showed that the amplitude of the P1 was larger for pretest compared to posttest and delay (both

$p_{Fisher} < .05$). When each of the SF conditions were analyzed in separate ANOVAs, the main effect of session was significant for HSF images, $F(2,30) = 3.82$, $p < .05$, with pretest > posttest = delay (both $p_{Fisher} < .05$) and marginal for AllSF, $F(2,30) = 3.17$, $p = .056$, and LSF, $F(2,30) = 3.22$, $p = .054$, images.

3.2.2.2. N170 component analysis. A plot of the amplitude of the N170 is presented in Fig. B. 3 and complete results of the statistical analysis are presented in Table B. 5. The main effect of SF, $F(2,30) = 19.06$, $p < .001$, showed that the amplitude of the N170 was smaller for AllSF compared to HSF and LSF images (both $p_{Fisher} < .001$). The main effect of session, $F(2,30) = 19.06$, $p < .001$, revealed that the amplitude of the N170 was increased by training (pretest < posttest = delay) (both $p_{Fisher} < .05$). The hemisphere \times session interaction, $F(2,30) = 3.49$, $p < .05$, indicated that the training effect was larger over the right hemisphere compared to the left hemisphere ($p_{Fisher} < .05$ 0.05 posttest and $p_{Fisher} < .001$ for delay session). When each of the SF conditions were analyzed in separate ANOVAs, the main effect of session was

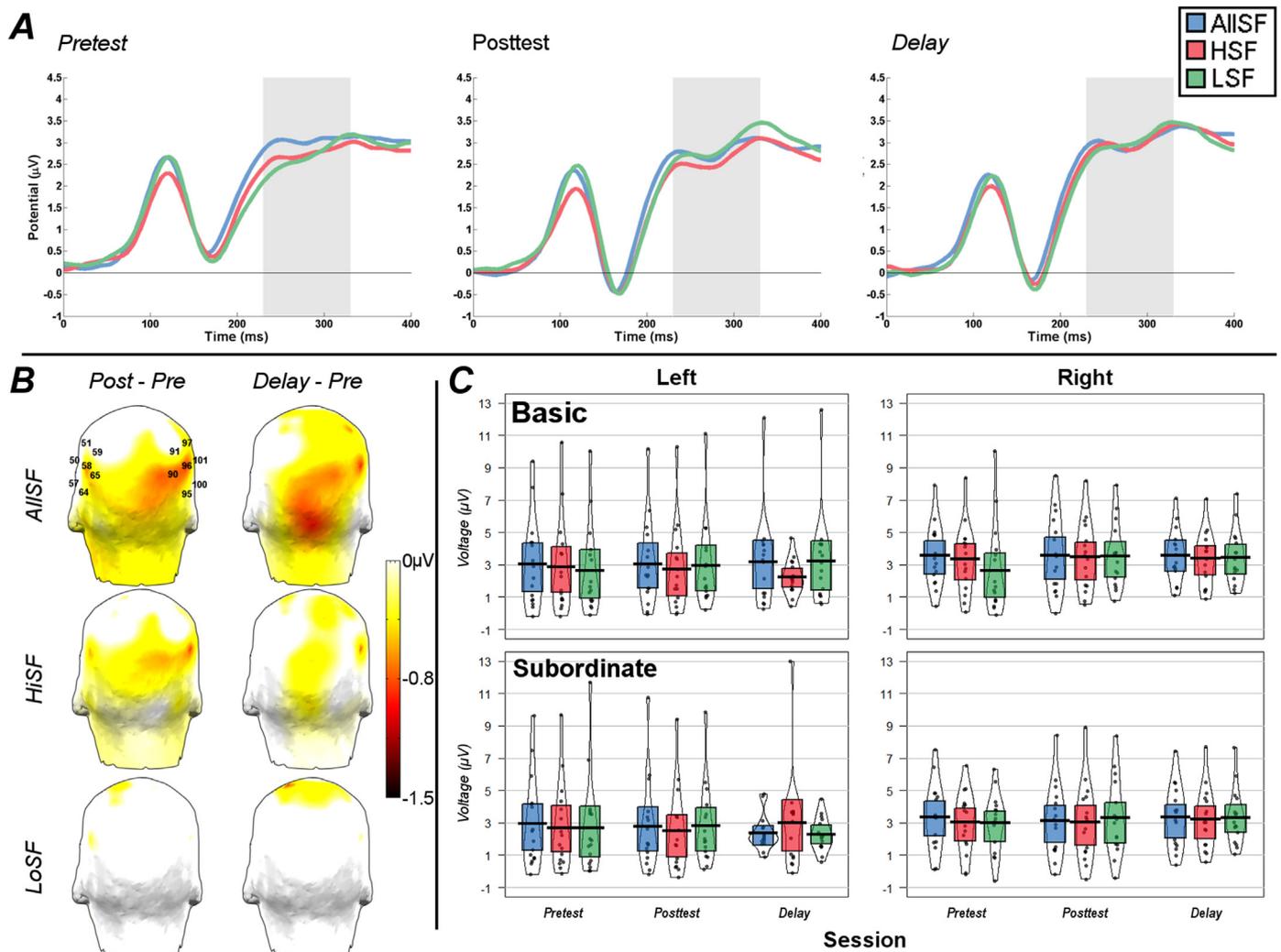


Fig. 8. **A.** N250 amplitude for pretest, posttest and delay, for AllSF, HSF and LSF images and for subordinate training only, collapsed over hemispheres, order and exemplar types. Gray regions show time windows for the N250 ERP component **B.** Topography differences between posttest and pretest, and delay and pretest, for AllSF, HSF and LSF images at time 280 ms **C.** N250 amplitude for pretest, posttest and delay, for AllSF, HSF and LSF images, collapsed over order and exemplar types. Colored regions represent the 95% confidence interval. Thick black lines represent the mean. Individual data points represent the mean amplitude for each participant.

significant for AllSF, $F(2,30) = 4.18, p < .05$, and HSF, $F(2,30) = 4.21, p < .05$, with pretest > posttest = delay (all $p_{Fisher} < .05$), but not for LSF conditions.

3.2.2.3. N250 component analysis. Fig. 8 presents the amplitude of the N250 for pretest, posttest and delay, for the three SF conditions. Complete results of the statistical analysis are presented in Table B. 6. The main effect of level, $F(1,15) = 19.26, p < .01$, indicated that the amplitude of the N250 was more negative for subordinate level compared to basic level training. The significant SF \times session interaction, $F(4,60) = 5.57, p < .02$, was followed up by: (1) SF effects within each session, and (2) training session effects within each SF. First, for pretest, the amplitude of the N250 was ordered LSF < HSF < AllSF (with LSF having the most negative amplitude, and hence largest N250) (all $p_{Fisher} < .05$). For posttest, the amplitude of the N250 was ordered HSF < LSF = AllSF (all $p_{Fisher} < .05$). No difference was observed between conditions during delay session. Second, for AllSF, no training effect was observed. For HSF, N250 amplitude was more negative for pretest and posttest than delay sessions (both $p_{Fisher} < .05$). For LSF, N250 amplitude was more negative for posttest and delay than pretest sessions (both $p_{Fisher} < .001$). As previously indicated, we have no hypotheses concerning order, so the last

significant interaction in Table B. 6 (SF \times order) is not discussed. When each of the SF conditions were analyzed in separate ANOVAs, no significant session effects were observed.

4. Discussion

The aim of the current study was to examine the role of color and spatial frequency (SF) on the early acquisition of perceptual expertise after one week of laboratory training with bird stimuli. Participants learned to categorize 10 species of finches and warblers at either the species or family levels. Birds were shown in their natural color during training, but we tested the effect of color by presenting birds in 3 conditions during pretest and two posttests (1-day posttest; 1-week delay): congruent color, incongruent color and grayscale. The effect of SF was tested by presenting birds in 3 conditions during pre- and posttests: all SF (AllSF), low SF (LSF < 8 cpi) and high SF (HSF > 8 cpi). We will discuss separately the effects of the color and SF manipulations.

4.1. Color

Both congruent and incongruent colors produced better performance (higher d' and shorter RT) than grayscale images independent of

training. All color conditions showed improved accuracy from training in both the posttest and the delay session, and improvement was strongest when color was congruent. Matching performance benefited more from subordinate- than basic-level training (replicating Scott et al., 2006, 2008; Jones et al., 2018). Prior to training, congruent and incongruent colors did not differ and both showed better matching discrimination than grayscale images. This result is consistent with the benefits of color for delineating shape and surface segmentation in early visual processing (Wurm et al., 1993; Gegenfurtner and Rieger, 2000). However, after subordinate-level training, participants performed best with congruent colors relative to the incongruent and grayscale images. These findings highlight the importance of color knowledge in subordinate-level identification and suggest that color congruency effects were dependent on learning. The results show that subordinate-level training with color images may be important for supporting the early acquisition of perceptual expertise. Previous work (Hagen et al., 2014) suggests that color knowledge is automatically applied by experts, at least when color is diagnostic for that domain. In Hagen et al.'s (2014) study, bird experts were asked to ignore color information, but this was hard for them, due to a recognition strategy in which color encoding is an implicit and automatized process. The authors proposed that color contributed to the recognition advantage even when structural information is sufficient for accurate recognition. This suggests that, for experts, object representations in their domain of expertise contain color information. In our study, we show that novices integrate color information into their object representation after minimal training, similar to what was observed for real-world experts. Our results partly differ from Jones et al.'s (2018) training study, which showed that accuracy was greater for color compared to grayscale but the color effect did not differ between pre- and post-test. The different pattern of results between our study and Jones et al.'s (2018) may be due to stimulus difference: they used images of computer-generated artificial objects and participants had no pre-experimental experience with these objects compared to birds used in this study.

The amplitude of P1 decreased with training, in contrast to previous work that did not report P1 training effects and did not appear to be present in the published waveform figures (Scott et al., 2006, 2008; Jones et al., 2018). The amplitude of P1 generally varies with the amount of attention (Mangun et al., 1993; Clark and Hillyard, 1996). The effect observed in our study could be explained by adults decreasing attention or habituation from pretest to the posttest and delay sessions. It is possible that P1 habituation was present in the current investigation and not previous studies (Scott et al., 2006, 2008; Jones et al., 2018) because here EEG was also recorded during the training sessions and not just at pre- and post-test, in which case participants may pay more attention if they think these sessions are more important. On the other hand, such general attention/habituation influences would not be expected to affect the P1 to congruent images more so than to incongruent and grayscale images, as we observed.

Results for the N170 and N250 components replicated previous work, which has generally showed performance increases to parallel N250 increases more than the N170 (Scott et al., 2006, 2008). Training increased N170 amplitude for both posttest and delay, following both basic- and subordinate-level training, for each color condition when analyzed separately. This result is consistent with previous studies using birds (Scott et al., 2006) and cars (Scott et al., 2008) but differs from Jones et al. (2018) who used computer-generated artificial objects and reported that N170 amplitude was more enhanced for subordinate-trained objects compared to basic-trained objects. As for accuracy, this may be explained by differences in participant's past experience with birds in general. In the present study, training increased N250 amplitude for both the posttest and the delay session, but with a larger increase after subordinate- relative to basic-level training. For each of the three color conditions, we observed a larger amplitude after subordinate- than basic-level training (posttest and delay session). Furthermore, subordinate-level training effects on N250 were larger for

congruent images compared to incongruent and grayscale images. These results differed from Jones et al. (2018) who did not find N250 differences between color and grayscale image. Again, this difference is probably due to participant previous experience with the objects used in the two different studies. Consistent with behavioral results, no N250 difference was observed between congruent and incongruent images until after subordinate-level training. These results show the importance of color information for subordinate-level discrimination, indexed by the N250. However, some findings were not entirely consistent between the accuracy and N250 results. One inconsistency was that the accuracy training effects generalized to incongruent and grayscale images after a week delay, whereas the N250 training effects only generalized to incongruent and grayscale images a day after training. Thus, the processes supporting accuracy improvements may be distinct from those that underlie the N250, or may include additional neural resources not measured by the N250. One possibility is that processes underlying the N170 contribute to the performance effects, as color and training did not interact at the level of the N170, and N170 training effects persisted for a week in all conditions.

Taken together, behavioral and ERP results showed how short-term training changed the way color was used behaviorally, and that this change was mirrored in neural correlates related to high-level vision. It is known that color plays a critical role in both low-level and high-level vision (Tanaka et al., 2001). At the lower level, color helps to segment objects from the background and different parts of the object. This perspective is supported by our pretest results showing similar accuracy and N250 effects on congruent and incongruent color images before training. In this case, participant's *a priori* knowledge of the "real" color of objects did not benefit their performance before training. This is in line with previous studies arguing that color is sensory in nature and not related to people's knowledge of the colors of things (Wurm et al., 1993). However, subordinate-level training, and by consequence learning, leads to a higher level of recognition where images were categorized more readily and elicited a larger N250 amplitude when presented in their congruent colors compared to incongruent. These results are in line with previous studies claiming that color is an intrinsic component of visual representations of objects and not only used to help segmentation or other precursors to object recognition (Naor-Raz et al., 2003). Thus, our posttest results support the role of color in high-level visual mechanisms. To summarize, our results are showing that color plays an important role at different levels of vision processing, with pre-training effects showing the effect of color on low-level vision and post-subordinate-training effects supporting the role of color in high-level vision.

Overall, our results are contrary to edge-based theories (Ostergaard and Davidoff, 1985; Biederman and Ju, 1988; Biederman and Gerhardstein, 1993) which argue that surface cues, like color, are generally less efficient for accessing object representations. The edge-based theories argue that color is not related to vision but rather to non-vision/semantic processes (Davidoff and Ostergaard, 1988). However, our results are in line with surface-plus-edge-based theories which claim that both shape contours and color contribute to recognition (Humphrey et al., 1994; Wurm et al., 1993). Thus, beyond its low-level contribution to edge segmentation, color plays an important role in high-level vision by providing diagnostic information that facilitates object recognition (Mapelli and Behrmann, 1997; Tanaka et al., 2001).

4.2. Spatial frequency

SF effects did not interact with training, but training improved performance for all image types. Because training images were always color congruent, it is important to keep in mind that all three SF conditions (AllSF, HSF, and LSF) represent generalization to manipulated images. Accuracy before as well as after training was higher for AllSF compared to HSF, and for AllSF and HSF compared to LSF. Reaction time was longer for LSF images compared to both HSF and AllSF

images. Faster answers were observed for basic- relative to subordinate-level training.

These results are similar to those of previous studies. Collin and McMullen (2005) used vehicles (planes, boats, cars) and animals (insects, birds, dogs) and demonstrated that the LSF condition selectively impaired subordinate- but not basic-level category verification. In Hagen et al.'s (2016), bird experts and novices demonstrated higher accuracy for birds presented in midrange SFs (8–32 cpi) compared to low- (2–4 and 4–8 cpi) and high- (32–64 cpi) range SF. These results are also partly consistent with Jones et al. (2018) who showed that subordinate matching accuracy with novel visual species was greater for AllSF and HSF relative to LSF. Current findings suggest that the effects of SF did not qualitatively change with learning in contrast to the changes observed for color. Rather, the SF manipulation magnified the qualitatively similar training effects for AllSF and HSF compared to LSF. However, we did expect reaction time differences between SF conditions to be larger after training than before training, based on Hagen et al.'s (2016) finding the SF impacts reaction time for experts but not novices. Perhaps our matching task was relatively insensitive to reaction time difference because speed was not emphasized to subjects, like it was in Hagen et al.'s (2016) category verification task. Our results are also consistent with previous studies exploring SF manipulations on object recognition using geometric figures (Boeschoten et al., 2005), airplanes (Harel and Bentin, 2009) or cars (Harel and Bentin, 2013), which suggest that basic-level object recognition relies on the general shapes of objects (e.g., external contour/global shape; presence of certain component features like beak for birds), whereas subordinate-level recognition relies more on specific diagnostic details (e.g., variations on a fixed set of features like shape of beak, wings, tail) contained in the higher spatial frequency bands. Our results partly support this interpretation, with small accuracy differences observed between the basic- and subordinate-level for LSF (i.e., general shapes of objects) but larger accuracy differences for subordinate- compared to basic-level training for AllSF. However, for HSF, we observed the same pattern as for LSF. This difference could be explained by the cutoff applied to stimuli in HSF which was higher in previous work (16 cpi in Collin and McMullen, 2005, 21 cpi in Boeschoten et al., 2005 and 54 cpi in Harel and Bentin, 2009).

The electrophysiological results show that P1, N170 and N250 were all affected by the SF manipulation, but there were no interactions between SF and training level or session. As for the color manipulation we observed a decrease of P1 amplitude with training. The P1 also showed smaller amplitude for HSF compared to AllSF and LSF. One previous study on object processing with high-pass and low-pass filtering did not report any effect on P1 (Goffaux et al., 2003), even though it is a component which is sensitive to the low-level visual properties determining the overall stimulus visibility. However, our results are consistent with previous studies showing that the P1 amplitude to faces was increased by low-pass filtering (Alorda et al., 2007; Nakashima et al., 2008).

The N170 exhibited a smaller amplitude for AllSF compared to HSF and LSF, and, when SF conditions were taken separately, we observed training effects for AllSF and HSF, where the amplitude of N170 was larger after training, but not for the LSF condition. N170 results replicated previous results showing that the N170 is mainly sensitive to the relatively familiarity of object categories with an increase of amplitude following both basic and subordinate training (Scott et al., 2006, 2008). The SF manipulation effect on N170 amplitude was also observed by Jones et al. (2018) with their novel visual objects. These results are opposite to previous work showing that the amplitude of N170 to cars was larger for low spatial frequencies (< 8 cpi) compared to all spatial frequencies and not different between all spatial frequencies and high spatial frequencies (> 32 cpi, Goffaux et al., 2003). Perhaps these differences reflect variation in the importance of different frequency ranges for different types of objects. This perspective is supported by studies showing that the spatial frequency characteristics

vary across different types of stimuli, as for example faces (Blickhan et al., 2011), natural scenes (Simoncelli and Olshausen, 2001) or art images (Redies, 2007). However, these results are generally consistent with findings suggesting that manipulations that impair categorization (e.g., face inversion) can result in larger N170 amplitude (Rossion and Jacques, 2011). Furthermore, we observed an increased right lateralized N170 component after training. This is consistent with previous work with faces (Rossion et al., 2003) or novel objects (Jones et al., 2018) who showed that subordinate-level training effects appear to be more right lateralized for the N170 compared to the N250 component.

As previously summarized along with the color results, training effects on the N250 were not significant for AllSF (i.e. grayscale) images. Thus, even our baseline condition for assessing SF effects did not show effects of training on the N250. The LSF images showed effects of training, with N250 amplitudes being more negative after training (posttest and delay) than before training. The HSF images showed an unusual pattern with N250 amplitudes being more negative for pretest and posttest compared to the delay session. None of the training effects observed on N250 interacted with training level, as it has been observed in previous training studies with color images of birds and cars (Scott et al., 2006, 2008) as well as for the congruent and incongruent color images in the present study. Thus, as previously observed for the N250 in the context of the color manipulation, training effects on the N250 did not generalize across the SF manipulation as readily as observed for discrimination performance.

5. Conclusions

This study examined how the effects of perceptual expertise training generalized to manipulations of color and spatial frequency (SF). Replicating previous training studies, performance benefited more from subordinate- than basic-level training. Color manipulations suggest that color can be learned, even through short-time training, when it is diagnostic for a category of objects. Before training, any color helps, but color congruence effects (congruent > incongruent) only emerged after subordinate-level training. SF manipulations influenced performance such that subordinate-level matching performance was better for images retaining AllSF than those containing only HSF, and HSF were better than LSF, but these SF effects did not interact with training.

The neural correlates of these effects were measured with ERPs. The N170 component was sensitive to SF manipulations, but not color. This result is in line with face literature which reported that shape is more relevant than color during earlier stages of processing, as mainly measured by the N170 component (Caharel et al., 2009; Itz et al., 2014). As mentioned previously, in this study we considered the SF manipulation to be predominantly a manipulation of shape. N170 SF effects did not interact with training, and training effects generalized to all manipulations except LSF. This somewhat broad generalization is consistent with the view that the N170 indexes basic-level categorization processes (Scott et al., 2006, 2008; Rossion and Jacques, 2011) that would be less sensitive to perceptual details. The N250 was sensitive to both color and SF manipulations. Subordinate level training effects on the N250 did not generalize to any of the non-color conditions. Like performance, the color ERP differences that were not present before training were demonstrated after training. Specifically, congruent and incongruent images did not differ before training, but amplitudes were more negative for congruent than incongruent images after subordinate-level training. These results suggest that color plays an important role in both low-level vision, supported by pretest results, and high-level vision, as shown by posttest results. Together, this supports surface-plus-edge-based theories for object processing and recognition (Wurm et al., 1993; Mapelli and Behrmann, 1997; Tanaka et al., 2001). The SF results indicated that shape information is also important for object recognition, again supporting surface-plus-edge-based theories, but that more extensive training is needed to use SF information more efficiently, as shown in Hagen et al. (2016) with real-expert bird

experts.

Acknowledgements

This work was supported by an Army Research Institute for the Behavioral and Social Sciences, USA contract W5J9CQ-11-C-0047 to L. Scott, T. Curran, and J. Tanaka. The views, opinions, and/or findings contained in this manuscript are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.neuropsychologia.2018.11.011](https://doi.org/10.1016/j.neuropsychologia.2018.11.011).

References

- Alorda, C., Serrano-Pedraza, I., Campos-Bueno, J.J., Sierra-Vázquez, V., Montoya, P., 2007. Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia* 45 (14), 3223–3233. <https://doi.org/10.1016/j.neuropsychologia.2007.06.017>.
- Bentin, S., McCarthy, G., Perez, E., Puce, A., Allison, T., 1996. Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* 8, 551–565. <https://doi.org/10.1162/jocn.1996.8.6.551>.
- Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94 (2), 115. <https://doi.org/10.1037/0033-295X.94.2.115>.
- Biederman, I., Gerhardstein, P.C., 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol.: Hum. Percept. Perform.* 19 (6), 1162. <https://doi.org/10.1037/0096-1523.19.6.1162>.
- Biederman, I., Ju, G., 1988. Surface versus edge-based determinants of visual recognition. *Cogn. Psychol.* 20 (1), 38–64. [https://doi.org/10.1016/0010-0285\(88\)90024-2](https://doi.org/10.1016/0010-0285(88)90024-2).
- Blickhan, M., Kaufmann, J.M., Denzler, J., Schweinberger, S.R., Redies, C., 2011. 1/fp Characteristics of the Fourier power spectrum affects ERP correlates of face learning and recognition. *Biol. Psychol.* 88 (2–3), 204–214. <https://doi.org/10.1016/j.biopsycho.2011.08.003>.
- Boeschoten, M.A., Kemner, C., Kenemans, J.L., Van Engeland, H., 2005. The relationship between local and global processing and the processing of high and low spatial frequencies studied by event-related potentials and source modeling. *Cogn. Brain Res.* 24 (2), 228–236. <https://doi.org/10.1016/j.cogbrainres.2005.01.021>.
- Busey, T.A., Vanderkolk, J.R., 2005. Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vis. Res.* 45 (4), 431–448. <https://doi.org/10.1016/j.visres.2004.08.021>.
- Caharel, S., Jiang, F., Blanz, V., Rossion, B., 2009. Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. *Neuroimage* 47 (4), 1809–1818. <https://doi.org/10.1016/j.neuroimage.2009.05.065>.
- Carmel, D., Bentin, S., 2002. Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 83 (1), 1–29. [https://doi.org/10.1016/S0010-0277\(01\)00162-7](https://doi.org/10.1016/S0010-0277(01)00162-7).
- Cavanagh, P., 1987. Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity, and shape. *Comput. Vis. Graph. Image Process.* 37 (2), 171–195. [https://doi.org/10.1016/0031-3203\(87\)80001-4](https://doi.org/10.1016/0031-3203(87)80001-4).
- Cheung, O.S., Gauthier, I., 2014. Visual appearance interacts with conceptual knowledge in object recognition. *Front. Psychol.* 5, 955. <https://doi.org/10.3389/fpsyg.2014.00793>.
- Clark, V.P., Hillyard, S.A., 1996. Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *J. Cogn. Neurosci.* 8 (5), 387–402. <https://doi.org/10.1162/jocn.1996.8.5.387>.
- Collin, C.A., McMullen, P.A., 2005. Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Atten. Percept. Psychophys.* 67 (2), 354–364. <https://doi.org/10.3758/BF03206498>.
- Curran, T., Tanaka, J.W., Weiskopf, D.M., 2002. An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2 (1), 1–18. <https://doi.org/10.3758/CABN.2.1.1>.
- Davidoff, J.B., Ostergaard, A.L., 1988. The role of colour in categorial judgements. *Q. J. Exp. Psychol.* 40 (3), 533–544. <https://doi.org/10.1080/02724988843000069>.
- Folstein, J.R., Monfared, S.S., Maravel, T., 2017. The effect of category learning on visual attention and visual representation. *Psychophysiology* 54 (12), 1855–1871. <https://doi.org/10.1111/psyp.12966>.
- Gauthier, I., Tarr, M.J., 1997. Becoming a “Greeble” expert: exploring mechanisms for face recognition. *Vis. Res.* 37 (12), 1673–1682.
- Gauthier, I., Behrmann, M., Tarr, M.J., 1999. Can face recognition really be dissociated from object recognition? *J. Cogn. Neurosci.* 11 (4), 349–370. <https://doi.org/10.1162/089992999563472>.
- Gauthier, I., Curran, T., Curby, K.M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. *Nat. Neurosci.* 6 (4), 428–432. <https://doi.org/10.1038/nn1029>.
- Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3 (2), 191–197. <https://doi.org/10.1038/72140>.
- Gegenfurtner, K.R., Rieger, J., 2000. Sensory and cognitive contributions of color to the recognition of natural scenes. *Curr. Biol.* 10 (13), 805–808. [https://doi.org/10.1016/S0960-9822\(00\)00563-7](https://doi.org/10.1016/S0960-9822(00)00563-7).
- Goffaux, V., Gauthier, I., Rossion, B., 2003. Spatial scale contribution to early visual differences between face and object processing. *Cogn. Brain Res.* 16 (3), 416–424. [https://doi.org/10.1016/S0926-6410\(03\)00056-9](https://doi.org/10.1016/S0926-6410(03)00056-9).
- Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* 7 (5), 555–562. <https://doi.org/10.1038/nn1224>.
- Hagen, S., Vuong, Q.C., Scott, L.S., Curran, T., Tanaka, J.W., 2014. The role of color in expert object recognition. *J. Vis.* 14 (9). <https://doi.org/10.1167/14.9.9>.
- Hagen, S., Vuong, Q.C., Scott, L.S., Curran, T., Tanaka, J.W., 2016. The role of spatial frequency in expert object recognition. *J. Exp. Psychol.: Hum. Percept. Perform.* <https://doi.org/10.1037/xhp0000139>.
- Harel, A., Bentin, S., 2013. Are all types of expertise created equal? Car experts use different spatial frequency scales for subordinate categorization of cars and faces. *PLoS One* 8 (6), e67024. <https://doi.org/10.1371/journal.pone.0067024>.
- Harel, A., Bentin, S., 2009. Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *J. Exp. Psychol.: Hum. Percept. Perform.* 35 (4), 1264. <https://doi.org/10.1037/a0013621>.
- Humphrey, G.K., Goodale, M.A., Jakobson, L.S., Servos, P., 1994. The role of surface information in object recognition: studies of a visual form agnostic and normal subjects. *Perception* 23 (12), 1457–1481. <https://doi.org/10.1068/p231457>.
- Itz, M.L., Schweinberger, S.R., Schulz, C., Kaufmann, J.M., 2014. Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. *NeuroImage* 102, 736–747. <https://doi.org/10.1016/j.neuroimage.2014.08.042>.
- Johnson, K.E., Mervis, C.B., 1997. Effects of varying levels of expertise on the basic level of categorization. *J. Exp. Psychol.: Gen.* 126 (3), 248. <https://doi.org/10.1037/0096-3445.126.3.248>.
- Jones, T., Hadley, H., Cataldo, A., Arnold, E., Curran, T., Tanaka, J., Scott, L., 2018. Neural and behavioral effects of subordinate-level training of novel objects across manipulations of color and spatial frequency. *Eur. J. Neurosci.* <https://doi.org/10.1111/ejn.13889>.
- Kaufmann, J.M., Schweinberger, S.R., Burton, A.M., 2009. N250 ERP correlates of the acquisition of face representations across different images. *J. Cogn. Neurosci.* 21 (4), 625–641. <https://doi.org/10.1162/jocn.2009.21080>.
- Limbach, K., Kaufmann, J.M., Wiese, H., Witte, O.W., Schweinberger, S.R., 2018. Enhancement of face-sensitive ERPs in older adults induced by face recognition training. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2018.08.010>.
- Mack, M.L., Wong, A.C., Gauthier, I., Tanaka, J.W., Palmeri, T.J., 2009. Time course of visual object categorization: fastest does not necessarily mean first. *Vision. Res.* 49 (15), 1961–1968. <https://doi.org/10.1016/j.visres.2009.05.005>.
- Mangun, G.R., Hillyard, S.A., Luck, S.J., 1993. IQ electrocortical substrates of visual selective attention. In: Meyer, D., Kornblum, S. (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*. 14. MIT Press, Cambridge, Massachusetts, pp. 219–243.
- Mapelli, D., Behrmann, M., 1997. The role of color in object recognition: evidence from visual agnosia. *Neurocase* 3 (4), 237–247. <https://doi.org/10.1080/13554799708405007>.
- Martens, F., Bulthé, J., van Vliet, C., de Beeck, H.O., 2018. Domain-general and domain-specific neural changes underlying visual expertise. *NeuroImage* 169, 80–93. <https://doi.org/10.1016/j.neuroimage.2017.12.013>.
- Nagai, J.I., Yokosawa, K., 2003. What regulates the surface color effect in object recognition: color diagnosticity or category. *Tech. Rep. Atten. Cogn.* 28, 1–4.
- Nakashima, T., Kaneko, K., Goto, Y., Abe, T., Mitsudo, T., Ogata, K., Tobimatsu, S., et al., 2008. Early ERP components differentially extract facial features: evidence for spatial frequency-and-contrast detectors. *Neurosci. Res.* 62 (4), 225–235. <https://doi.org/10.1016/j.neures.2008.08.009>.
- Naor-Raz, G., Tarr, M.J., Kersten, D., 2003. Is color an intrinsic property of object representation? *Perception* 32 (6), 667–680. <https://doi.org/10.1068/p5050>.
- Oliva, A., Schyns, P.G., 1997. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* 34, 72–107. <https://doi.org/10.1006/cogp.1997.0667>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1. <https://doi.org/10.1155/2011/156869>.
- Ostergaard, A.L., Davidoff, J.B., 1985. Some effects of color on naming and recognition of objects. *J. Exp. Psychol.: Learn. Mem. Cogn.* 11 (3), 579. <https://doi.org/10.1037/0278-7393.11.3.579>.
- Pierce, L.J., Scott, L., Boddington, S., Droucker, D., Curran, T., Tanaka, J., 2011. The n250 brain potential to personally familiar and newly learned faces and objects. *Front. Hum. Neurosci.* 5 (111). <https://doi.org/10.3389/fnhum.2011.00111>.
- Redies, C., 2007. A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spat. Vis.* 21 (1), 97–117. <https://doi.org/10.1163/156856807782753886>.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. *Cogn. Psychol.* 8 (3), 382–439. [https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X).
- Rossion, B., Jacques, C., 2011. The N170: understanding the time-course of face perception in the human brain. *The Oxford Handbook of ERP Components*. pp. 115–142. <https://doi.org/10.1093/oxfordhb/9780195374148.013.0064>.
- Rossion, B., Collins, D., Goffaux, V., Curran, T., 2007. Long-term expertise with artificial objects increases visual competition with early face categorization processes. *J. Cogn. Neurosci.* 19 (3), 543–555. <https://doi.org/10.1162/jocn.2007.19.3.543>.
- Rossion, B., Gauthier, I., Goffaux, V., Tarr, M.J., Crommelinck, M., 2002. Expertise

- training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol. Sci.* 13 (3), 250–257. <https://doi.org/10.1111/1467-9280.00446>.
- Rossion, B., Joyce, C.A., Cottrell, G.W., Tarr, M.J., 2003. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage* 20 (3), 1609–1624. <https://doi.org/10.1016/j.neuroimage.2003.07.010>.
- Schweinberger, S.R., Huddy, V., Burton, A.M., 2004. N250r: a face-selective brain response to stimulus repetitions. *Neuroreport* 15 (9), 1501–1505. <https://doi.org/10.1097/01.wnr.0000131675.00319.42>.
- Schweinberger, S.R., Pickering, E.C., Jentzsch, I., Burton, A.M., Kaufmann, J.M., 2002. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cogn. Brain Res.* 14 (3), 398–409. [https://doi.org/10.1016/S0926-6410\(02\)00142-8](https://doi.org/10.1016/S0926-6410(02)00142-8).
- Scott, L., Tanaka, J., Sheinberg, D., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. *J. Cogn. Neurosci.* 18 (9), 1453–1465. <https://doi.org/10.1162/jocn.2006.18.9.1453>.
- Scott, L., Tanaka, J., Sheinberg, D., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. *Brain Res.* 1210, 204–215. <https://doi.org/10.1016/j.brainres.2008.02.054>.
- Simoncelli, E.P., Olshausen, B.A., 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24 (1), 1193–1216. <https://doi.org/10.1146/annurev.neuro.24.1.1193>.
- Stein, T., Reeder, R.R., Peelen, M.V., 2016. Privileged access to awareness for faces and objects of expertise. *J. Exp. Psychol.: Hum. Percept. Perform.* 42 (6), 788. <https://doi.org/10.1037/xhp0000188>.
- Tanaka, J.W., Curran, T., 2001. A neural basis for expert object recognition. *Psychol. Sci.* 12 (1), 43–47. <https://doi.org/10.1111/1467-9280.00308>.
- Tanaka, J.W., Pierce, L.J., 2009. The neural plasticity of other-race face recognition. *Cogn. Affect. Behav. Neurosci.* 9 (1), 122–131. <https://doi.org/10.3758/CABN.9.1.122>.
- Tanaka, J.W., Presnell, L.M., 1999. Color diagnosticity in object recognition. *Percept. Psychophys.* 61 (6), 1140–1153. <https://doi.org/10.3758/BF03207619>.
- Tanaka, J.W., Taylor, M., 1991. Object categories and expertise: is the basic level in the eye of the beholder? *Cogn. Psychol.* 23 (3), 457–482. [https://doi.org/10.1016/0010-0285\(91\)90016-H](https://doi.org/10.1016/0010-0285(91)90016-H).
- Tanaka, J.W., Curran, T., Sheinberg, D.L., 2005. The training and transfer of real-world perceptual expertise. *Psychol. Sci.* 16 (2), 145–151. <https://doi.org/10.1111/j.0956-7976.2005.00795.x>.
- Tanaka, J.W., Curran, T., Porterfield, A.L., Collins, D., 2006. Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *J. Cogn. Neurosci.* 18 (9), 1488–1497. <https://doi.org/10.1162/jocn.2006.18.9.1488>.
- Tanaka, J.W., Weiskopf, D., Williams, P., 2001. The role of color in high-level vision. *Trends Cogn. Sci.* 5 (5), 211–215. [https://doi.org/10.1016/S1364-6613\(00\)01626-0](https://doi.org/10.1016/S1364-6613(00)01626-0).
- Tucker, D.M., 1993. Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* 87 (3), 154–163. [https://doi.org/10.1016/0013-4694\(93\)90121-B](https://doi.org/10.1016/0013-4694(93)90121-B).
- Wurm, L.H., Legge, G.E., Isenberg, L.M., Luebker, A., 1993. Color improves object recognition in normal and low vision. *J. Exp. Psychol.: Hum. Percept. Perform.* 19 (4), 899. <https://doi.org/10.1037/0096-1523.19.4.899>.