The Role of Spatial Frequency in Expert Object Recognition

Simen Hagen
University of Victoria

Quoc C. Vuong
Newcastle University

Lisa S. Scott
University of Massachusetts Amherst

Tim Curran
University of Colorado Boulder

James W. Tanaka
University of Victoria

Novices recognize objects at the basic-category level (e.g., dog, chair, and bird) at which identification is based on the global form of the objects (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In contrast, experts recognize objects within their domain of expertise at the subordinate level (e.g., Sparrow or Finch) for which the internal object information may play an important role in identification (Tanaka & Taylor, 1991). To investigate whether expert recognition relies on internal object information, we band-pass filtered bird images over a range of spatial frequencies (SF) and then masked the filtered image to preserve its global form. In Experiment 1, bird experts categorized common birds at the family level (e.g., Robin or Sparrow) more quickly and more accurately than novices. Both experts and novices were more accurate when bird images contained the internal information represented by a middle range of SFs, and this finding was characterized by a quadratic function in which accuracy decreased toward each end of the SF spectrum. However, the experts, but not the novices, showed a similar quadratic relationship between response times and SF range. In Experiment 2, experts categorized Warblers and Finches at the more specific, species level (e.g., Wilson's Warbler or House Finch). Recognition was again fastest and most accurate for images filtered in the middle range of SFs. Collectively, these results indicate that a midrange of SFs contain crucial information for subordinate recognition, and that extensive perceptual experience can influence the efficiency with which this information is utilized.

Keywords: expert object recognition, real-world expertise, spatial frequency, object categories, visual features

Supplemental materials: http://dx.doi.org/10.1037/xhp0000139.supp

Color, motion, and form all play a critical role in early vision by, for example, segregating an object from the background (Livingstone & Hubel, 1987). These dimensions also contribute to later stages of object recognition. For example, objects are more quickly identified when shown in a congruent color than in an incongruent color (Rossion & Pourtois, 2001; Tanaka & Presnell, 1999) and biological movements (e.g., of humans) can be recognized based on motion information when form cues are highly degraded (Johansson, 1973). Although it has been shown that experience with objects can modulate the role that color plays in recognition (Hagen, Vuong, Scott, Curran, & Tanaka, 2014), we know little about how experience can modulate other dimensions such as form information. In this article, we tested the importance of internal form information and how its role in object recognition may change as function of perceptual experience.

Most objects are initially recognized at a level of categorization called the *basic* level (e.g., dogs, birds, and cars). For example, the basic-level category (e.g., bird) is the level at which objects are typically labeled and fastest to be recognized (Mack, Wong, Gauthier, Tanaka, & Palmeri, 2009; Rosch et al., 1976). At the basic level, the objects' global form (e.g., external contour) minimizes variation in within-category members (e.g., dogs), and maximizes variation across between-category members (e.g., birds vs. dogs vs. cars; Rosch et al., 1976). Thus, objects that are recognized at the basic level are maximally distinctive from one another with

respect to their global form. In contrast, objects at a subordinate level (e.g., Robin is a subordinate level of bird) show greater within-category similarity than basic-level categories, but less between-category variability. Thus, the global form of members of subordinate-level categories is more homogenous and demands additional perceptual encoding time to recognize on the basis of form information (Grill-Spector & Kanwisher, 2005; Joliecour, Gluck, & Kosslyn, 1984; Mack & Palmeri, 2011; Mack et al., 2009; Rosch et al., 1976).

The level that is initially accessed in recognition is not absolute, but is adaptable to experience. Natural experts, such as car, bird and dog experts, recognize objects in their domain of expertise equally fast at the subordinate and basic levels (Johnson & Mervis, 1997; Tanaka & Taylor, 1991). In the current study, we examine whether perceptual experience influences the degree to which internal form information is used for fast and accurate subordinate-level identification. In a recent study, Hagen et al. (2014) demonstrated that color played a larger role in object recognition for experts relative to novices. In this study, expert birdwatchers and novice participants were asked to categorize common birds (e.g., Robin, Sparrow) presented in their natural congruent color, non-natural incongruent color and grayscale. Unlike novices who utilized color only when they were relatively slow at responding, experts utilized color even on the fastest trials suggesting that they relied on representations in which color was readily available. Thus, the ability to quickly recognize objects at the subordinate level seem to be facilitated by internal representations in which internal object cues (e.g., color) are encoded to increase within-category variation when global form cues become less informative (e.g., trying to recognize birds from the same species).

Following our previous work (Hagen et al., 2014), here we examine the role that experience plays in forming representations that encode internal form features of an object in visual memory (e.g., feathers of a bird). Similar to Hagen et al. (2014), we chose the bird domain for two main reasons. First, bird-watching requires quick and accurate recognition of visually homogenous objects (in terms of their external contour) at subordinate species levels (e.g., Field Sparrow) for which internal features may play a critical role in making within-category identifications. Second, experienced bird watchers readily report internal details (e.g., patch around the eye) in feature listing tasks (Tanaka & Taylor, 1991).

Form information, such as internal details of an object, can be investigated by filtering the spatial frequency (SF) content of images. That is, an image can be decomposed into different SF components that describe changes in luminance information in the image and that provide the building blocks for complex forms. The fast variations in luminance captured by high spatial frequencies (HSFs) represent fine details and edges that may define internal object features. By comparison, the slower luminance variations captured by low spatial frequencies (LSFs) represent coarse details and their configurations, as well as the global form of the object (see Morrsion & Schyns, 2001).

Different regions of the SF spectrum contain different visual information that are important for recognition of objects. More importantly, previous studies show that different regions of the SF spectrum are used differently depending on the recognition task. First, different narrow ranges of the SF spectrum can contain the crucial information for recognition of different object categories, as shown with recognition of letters (Solomon & Pelli, 1994),

geometrical shape objects (Braje, Tjan, & Legge, 1995), and faces (Costen, Parker, & Craw, 1994, 1996; Fiorentini et al., 1983; Hayes, Morrone, & Burr, 1986; Näsänen, 1999; Tieger & Ganz, 1979). For example, the 8–16 cpi range has been repeatedly shown to be crucial for face recognition (e.g., Costen et al., 1994, 1996), whereas the bandwidth critical to letters is centered at about 3 cycles a letter (e.g., Solomon & Pelli, 1994). Second, different category levels (e.g., basic, subordinate) rely on information contained in different regions of the SF spectrum. For example, Collin and McMullen (2005) showed that subordinate-level recognition of nonface objects relied disproportionately on HSFs (>16 cycles/object) relative to basic-level recognition of the same objects. Third, different face recognition mechanisms have been linked to different SFs. Whereas recognition based on features is facilitated by HSFs, recognition based on the spatial relations of those features is facilitated by LSFs (Costen, Parker, & Craw, 1994, 1996; Fiorentini, Maffei, & Sandini, 1983; Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Sergent, 1986). Fourth, expert recognition of faces and cars rely on information contained in different regions of the SF spectrum (Harel & Bentin, 2013), indicating that expertise for different domains may rely on different SF information (also see Harel & Bentin, 2009). Collectively, these studies support the notion that observers' experience with the object domain may influence how SF information contributes to recognition.

In the current study, we investigated whether extensive experience with birds modulated the representation of internal features of birds (as captured by different SF ranges). To test this question, individuals with and without extensive bird experience were asked to categorize images of common birds at the family level (e.g., Sparrow, Robin, and Wren; Experiment 1) and at the species level (e.g., Field Sparrow, American Goldfinch; Experiment 2). We systematically degraded the birds' internal information while holding their global form constant by filtering the bird images over a range of SF bandwidths corresponding to 2–4 cycles per image (cpi), 4–8 cpi, 8–16 cpi, 16–32 cpi, and 32–64 cpi and masking these filtered images with the birds' contour. Any difference in performance between the SF conditions must be because of the internal object information conveyed by that SF range because global form was controlled for in each condition. The LSFs preserved very little internal details because of the spatial blurring, whereas the HSFs preserved fine internal features (see Figure 1). The midrange SFs, typically defined between 8 and 32 cpi (e.g., Hughes, Nozawa, & Kitterle, 1996), contained a mixture of internal features and their configurations. In line with studies suggesting that a midrange of SFs is important for recognition (e.g., Tjan, Braje, Legge, & Kersten, 1995; Näsänen, 1999), we hypothesized that expert and novice recognition will be facilitated by a midrange in which the crucial object features are contained. Moreover, we hypothesized that experience would increase the sensitivity to the optimal SF information range and moderate how efficiently that information is processed.

Experiment 1

In Experiment 1 the effects of spatial frequency filtering on family-level categorization of birds were assessed with bird experts and novices. The two groups were tested in a category verification task in which participants judged whether a category
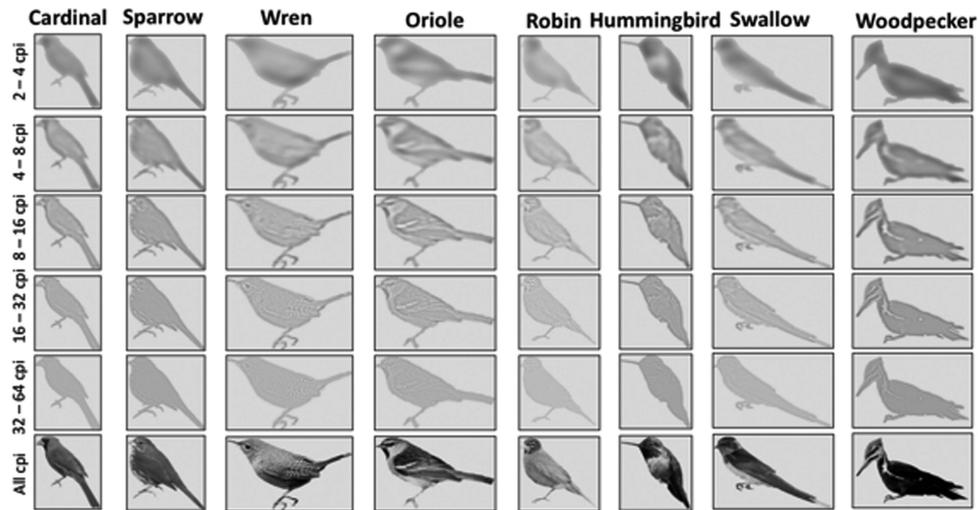
Figure 1. An example of stimuli filtered for 2–4, 4–8, 8–16, 16–32, and 32–64 cycles per image (cpi), and masked by the bird's contour. Note the systematic degradation in the frequency domain of the bird's internal features while its global form remains constant.

label and a subsequently presented bird image matched or not. Based on previous work (e.g., Hagen et al., 2014; Tanaka & Taylor, 1991), we predicted that bird experts would be faster and more accurate when categorizing the birds relative to the novices. Moreover, we predicted that experts would be more sensitive than novices to SF information contained within a midrange of SFs (e.g., Braje et al., 1995). As in Hagen et al. (2014), we performed an analysis of the response time (RT) distributions for both Experiment 1 and 2 to examine differences in the timing with which different ranges of SFs influence performance. If the internal information within the midrange of SFs is crucial for accurate and fast recognition, it should influence performance even on the fastest trials.

Method

Participants. Fifteen expert participants, ranging in age from 22–63 years (4 females, $M = 31.27$, $SD = 12.89$) were selected based on nominations from their bird-watching peers. Fifteen additional age- and education-matched (21–54 years of age) participants were selected to serve as the novice control group (7 females; $M = 30.20$, $SD = 8.35$). The novice participants had no prior experience in bird watching. However, seven out of the 15 novice participants had participated in a previous experiment on bird recognition (Hagen et al., 2014).[1] For the experts, nine out of the 15 participants had taken part in a previous study on bird recognition (Hagen et al., 2014). To assess the level of bird expertise in our participants, we used a bird expertise test (Hagen et al., 2014) in which participants judged whether two sequentially presented bird images belonged to the same or different species. The experts obtained a higher discrimination score ($d' = 2.15$) compared to the novices ($d' = 0.79$), $t(28) = 6.86$, $p < .001$. The data from two additional expert participants were lost because of technical issues, and the data from one additional novice were excluded because of poor performance (i.e., below 50% accuracy for the majority of bird families in the full spectrum condition).

Participants received monetary compensation for their participation.

Stimuli. Three images for each of eight common bird species (Cardinal, Oriole, Hummingbird, Robin, Sparrow, Swallow, Woodpecker, and Wren) were collected in part from the Internet and from an existing bird dataset (Wahlheim, Teune, & Jacoby, 2011) for a total of 24 bird images. The birds selected were among the 20 most frequently mentioned birds in a category norms study by Battig and Montague (1969).

Using Matlab, the bird images were band-pass filtered for different ranges of spatial frequencies (2–4, 4–8, 8–16, 16–32, and 32–64 cpi). The images were first Fourier transformed into the frequency domain. These transformed images were then multiplied with Gaussian filters with cutoff frequencies in each range (e.g., for the 8–16 cpi range, a low-pass cutoff of 8 cpi and a high-pass cutoff of 16 cpi). The filtered images were then inverse Fourier transformed back into the spatial domain. A mask of the external bird contour was applied such that the external contour remained constant across the spatial frequency manipulations. The masking necessarily distorts the frequency content of the band-passed filtered images. However, the important aspects were that the internal features were systematically degraded and that the global form (i.e., contour) was preserved. Each bird image was filtered in each frequency band condition, as illustrated in Figure 1. Images were cropped and scaled to fit within a frame of approximately 250 × 250 pixels and pasted on a gray background using Adobe Photoshop CS4. Images subtended a visual angle of approximately 4.09 degrees vertically and 5.03 degrees horizontally.

Procedure. Participants were tested in a category verification task. At the beginning of each trial, a ready prompt (i.e., "Get Ready") was displayed for 1.0 s before it was replaced by a category label (e.g., "Robin"). After 2.5 s, the category label was

---

[1] Analyses were done with and without these participants and the general pattern of effects did not change.

replaced by an image of a bird that remained on the screen until the participant responded. If the label and the image matched (e.g., the label "Robin" was followed by an image of a Robin), the participant was instructed to press the button on a keyboard labeled YES ("m" on the keyboard). If the label and the image mismatched (e.g., the label "Robin" was followed by an image of a Cardinal), the participant was instructed to press the button labeled NO ("c" on the keyboard). Before the task started, the participants were told which birds they would see in the experiment and instructed to respond as quickly and as accurately as possible.

The name foil for each target image was selected from the remaining seven bird families. Thus, the only labels that could appear in the experiment were the following: "Cardinal," "Oriole," "Wren," "Robin," "Hummingbird," "Woodpecker," "Swallow," and "Sparrow." Each bird name appeared an equal amount of times as a foil. Moreover, each bird was used as a foil and a correct label an equal amount of times.

Each bird exemplar (e.g., robin1) was displayed once in YES trial and once in a NO trial in each of the six SF conditions (2–4, 4–8, 8–16, 16–32, 32–64, All cpi). The orientation of each bird image was reversed across the horizontal axis for half of the YES trials and half of the NO trials. Consequently, each exemplar (e.g., cardinal1) would be presented six times facing left (three YES trials and three NO trials) and six times facing right (three YES trials and three NO trials). Three blocks were created to reduce the occurrence of the same bird exemplar in different SF conditions close in time (i.e., minimize image familiarity), and the order of these three blocks was counterbalanced across participants. Each block consisted of 96 trials (8 bird families, 3 exemplars, 2 facing directions, and 2 types of trial), for a total of 288 trials.

Results

Trials with RT 3 *SD* greater than each participant's grand mean were excluded from the analysis. Category items that were miscategorized on 50% of the trials in the All SF condition were also excluded from analysis. In total, three bird families were excluded across four novice participants (one Robin, two Orioles, and one Sparrow). Excluded trials based on these requirements included 120 (14 trials lost because of technical difficulties) for the novices and 49 for the experts out of a total of 2,160 trials for each group.

Similar to previous studies (e.g., Collin & McMullen, 2005; Mack, Wong, Gauthier, Tanaka, & Palmeri, 2009), all subsequent analyses were conducted using only the YES trials. The YES trials provide a measure of the extent that the category representation triggered by the label match with the image on the screen. For example, if the representation and the image have exactly the same content, then recognition should be faster than if the representation and the image have slightly different content. In our study, we apply the assumptions of the category verification paradigm to examine whether observer's representation includes specific spatial frequencies. In contrast, for the NO trials, the label triggers a category level representation (e.g., family level: Robin) that is not congruent with the bird image on the screen (e.g., Hummingbird). Thus, the label does not activate a bird representation that match with the bird in the image, and therefore, the label does not control the level at which the object is recognized and subsequently rejected. We do not know what category levels are used in the NO trials, and so we cannot make clear conclusions about the SF

content of family level (Experiment 1) and species level (Experiment 2) from these trials. See online supplementary material for analysis of the NO trials.

Accuracy. Figure 2 presents the mean accuracy data for YES trials as a function of SF condition and group. The accuracy data for experts and novices were analyzed in a mixed-design analysis of variance (ANOVA) with SF condition as a within-subjects factor and group as a between-subjects factor. The significant main effect of group, $F(1, 28) = 48.43$, $p < .001$, $\eta_p^2 = .63$, showed that the experts were more accurate than the novices. The main effect of SF condition was significant, $F(5, 140) = 15.75$, $p < .001$, $\eta_p^2 = .36$. In addition, as narrow ranges in the midrange SFs lead to best recognition of objects (Tjan et al., 1995), letters (Solomon & Pelli, 1994), and faces (e.g., Costen et al., 1994, 1996), we tested whether there was a quadratic trend in the accuracy and RT data as a function of SF condition (the All SF condition was removed because it included all SF ranges). There was a significant quadratic trend for both experts, $F(1, 14) = 5.20$, $p = .039$, $\eta_p^2 = .27$, and novices, $F(1, 14) = 6.83$, $p = .020$, $\eta_p^2 = .33$. Thus, both groups were more accurate when middle SF ranges were available in the images. Group and SF condition did not interact, $F(5, 140) = 0.76$, $p = .58$, nor was there a significant quadratic interaction between group and SF condition, $F(1, 28) = 0.52$, $p = .478$. Thus, SF filtering influenced experts and novices in the same way. Post-hoc comparisons aimed to evaluate the performance with the filtered images relative to the full-spectrum images (All SF condition) revealed that both experts and novices were more accurate in the All SF condition than all the other conditions (all ps $< .003$). For the post-hoc test in this and subsequent analyses, we report uncorrected p-values.

Response time. Figure 3 presents the mean RTs for correct YES trials as a function of SF condition and group. The RT data were analyzed in a mixed-design ANOVA using SF condition as a within-subjects factor and group as a between-subjects factor. The main effect of group, $F(1, 28) = 25.87$, $p < .001$, $\eta_p^2 = .48$, indicated that the experts were faster than the novices. The main effect of SF condition, $F(5, 140) = 6.90$, $p < .001$. $\eta_p^2 = .20$, demonstrated that the SF filtering influenced RT.
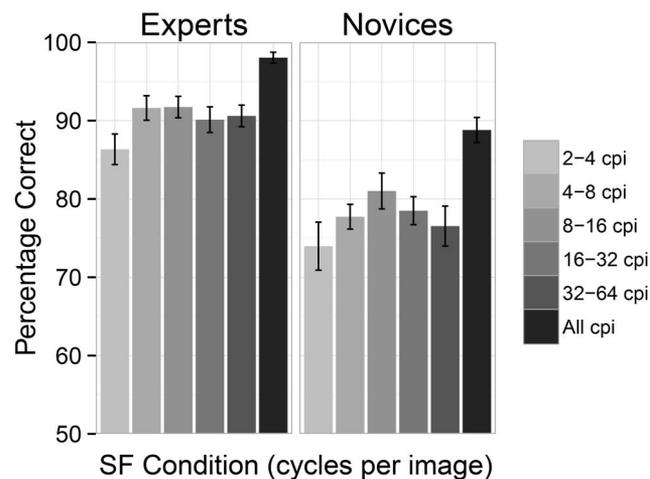

Figure 2. Experiment 1: Accuracy for each group (expert, novice) as a function of spatial frequency (SF) condition (2–4, 4–8, 8–16, 16–32, 32–64, and All cpi [cycles per image]). Error bars represent the *SEM*s.
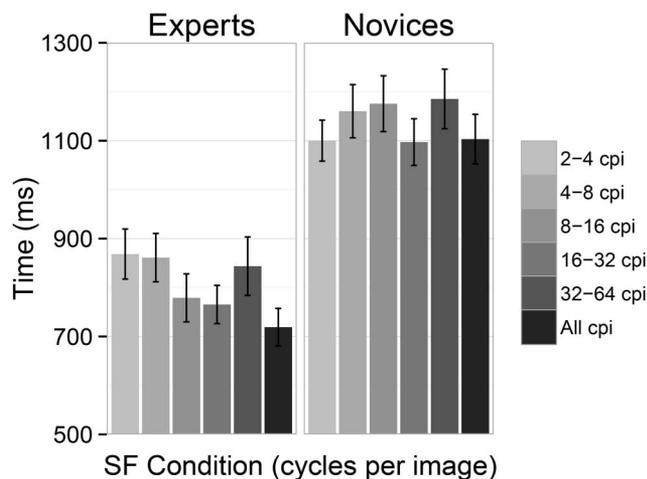
Figure 3. Experiment 1: Response time for each group (expert, novice) as a function of spatial frequency (SF) condition (2–4 cpi, 4–8 cpi, 8–16 cpi, 16–32 cpi, 32–64 cpi, and All cpi [cycles per image]). Error bars represent the *SEM*s.

In contrast to accuracy, SF condition interacted with group, $F(5, 140) = 3.54$, $p = .005$, $\eta_p^2 = .11$. More importantly, there was a significant quadratic trend in the interaction between group and SF condition, $F(1, 28) = 6.04$, $p = .020$, $\eta_p^2 = .18$. To follow up this interaction, we carried out separate ANOVAs for the novice and expert groups with SF condition as a within-subjects factor. For the novices, the main effect of SF condition was significant, $F(5, 70) = 2.59$, $p = .033$, $\eta_p^2 = .16$. However, unlike the accuracy, the trend analysis in which the All SF condition was excluded revealed that there was no significant quadratic trend, $F(1, 14) = 0.19$, $p = .673$. Post-hoc comparisons revealed that participants responded more quickly in the All SF condition than the 8–16 cpi ($p = .019$) and 32–64 cpi ($p = .040$) conditions, but not the remaining SF conditions (all $ps > .1$). To rule out the possibility of a quadratic trend if the 2–4 cpi condition was excluded (because of speed–accuracy trade-off), we ran the analysis without the 2–4 cpi condition. The results revealed no quadratic trend for the novices ($p = .176$). For the bird experts, the main effect of SF condition was significant, $F(5, 70) = 9.85$, $p < .001$, $\eta_p^2 = .41$. In contrast to the novices, there was a significant quadratic trend, $F(1, 14) = 11.30$, $p = .005$, $\eta_p^2 = .45$, indicating that participants responded more quickly with images containing middle SF ranges. Post-hoc comparisons revealed that participants responded more quickly in the All SF condition than all the other conditions (all $ps < .05$).

RT distribution analysis. Following our previous study with color (Hagen et al., 2014), we analyzed the RT distribution to investigate when the quadratic trend observed in the RT analysis for experts occurred and whether similar trends occurred for novices at different time points. In our earlier work, we combined accuracy and RTs into a single inverse efficiency measure (Hagen et al., 2014). Here we analyzed RTs directly as accuracy was relatively low in this study (<90%) compared with our previous study (making the inverse efficiency measure less robust; Bruyer & Brysbaert, 2011) and because speed–accuracy trade-offs did not complicate the interpretation of the results in this study.

To examine the distribution of correct RT, each participant's trials were ranked from the fastest to the slowest within each SF

condition before they were grouped into four bins containing the fastest 25% of the responses (i.e., quartile bin 1), the next 25% of responses (i.e., quartile bin 2), and so on. Within each bin, the average correct RT for each SF condition was calculated.

The data were first analyzed in a mixed-design ANOVA using SF condition and bin as within-subjects factors, and group as a between-subjects factor. The main effects of bin, $F(3, 84) = 283.15$, $p < .001$, $\eta_p^2 = .91$, SF condition, $F(5, 140) = 5.59$, $p < .001$, $\eta_p^2 = .17$, and group, $F(1, 28) = 24.68$, $p < .001$, $\eta_p^2 = .47$, were significant. The two-way interactions between SF condition and bin, $F(15, 420) = 3.99$, $p < .001$, $\eta_p^2 = .13$, between bin and group, $F(3, 84) = 19.30$, $p < .001$, $\eta_p^2 = .41$, and between group and SF condition, $F(4, 140) = 3.63$, $p = .004$, $\eta_p^2 = .12$, were significant. Importantly, the three-way interaction between bin, SF condition, and group was significant, $F(15, 420) = 2.22$, $p = .005$, $\eta_p^2 = .7$. Furthermore, there was a significant quadratic trend in this three-way interaction, $F(1, 28) = 7.39$, $p = .011$, $\eta_p^2 = .21$.

To better understand the three-way interaction, the groups were independently analyzed in a repeated measures ANOVA using SF condition and bin as within-subjects factors. Figure 4 presents the mean RTs for correct YES trials as a function of bin, SF condition, and group. For the novices, the main effect of bin, $F(3, 42) = 151.16$, $p < .001$, $\eta_p^2 = .92$) was significant. However, the main effect of SF condition, $F(5, 70) = 2.06$, $p = .081$, and the interaction between these two variables, $F(15, 210) = 1.35$, $p = .175$, were not significant. Moreover, a trend analysis at each bin showed that there was no significant quadratic trend for any bins (all $ps > .17$). Post-hoc comparisons revealed that participants responded more quickly in the All SF condition than the 8–16 cpi ($p = .024$) and the 32–64 cpi ($p = .026$) conditions in bin 3. They responded more quickly in the All SF condition than the 8–16 cpi ($p = .048$) condition in bin 4. These findings are consistent with the results of the RT analysis for novices above.

In contrast to novices, for the bird experts, the main effects of bin, $F(3, 42) = 151.40$, $p < .001$, $\eta_p^2 = .92$, and SF condition, $F(5, 70) = 9.08$, $p < .001$, $\eta_p^2 = .39$, were significant. Moreover, SF condition interacted with bin, $F(15, 210) = 7.84$, $p < .001$, $\eta_p^2 = .36$. Trend analysis revealed a quadratic trend in each bin similar to the trend we found for the RT analysis (bin 1: $F(1, 14) = 9.50$, $p = .008$, $\eta_p^2 = .40$; bin 2: $F(1, 14) = 8.04$, $p = .013$, $\eta_p^2 = .37$; bin 3: $F(1, 14) = 6.81$, $p = .021$, $\eta_p^2 = .33$; bin 4: $F(1, 14) = 7.37$,
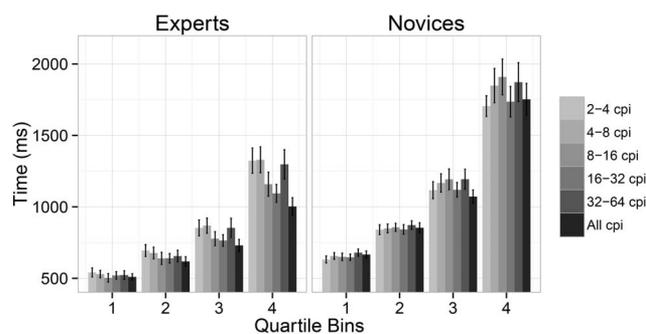

Figure 4. Experiment 1. Distribution of response times for the experts and novices. Bin 1 contains the 25% fastest responses of each participant. Bin 2 contains the next 25% fastest responses, and so on. Error bars represent the *SEM*s.

$p = .017$, $\eta_p^2 = .35$). Post-hoc comparisons revealed that in bin 2, participants responded more quickly in the All SF condition than the low SF conditions (2–4 and 4–8 cpi, ps $<$ .006). In bin 3, they responded more quickly in the All SF than the low SF (2–4 and 4–8 cpi, ps $<$ .002), 8–16 cpi ($p = .019$), and high SF (32–64 cpi, $p = .017$) conditions. In bin 4, they responded more quickly in the All SF condition than all the other conditions (all ps $<$ .05).

The main finding of Experiment 1 was that SF filtering differentially affected RTs of bird experts and bird novices whereas the effect of filtering on accuracy was the same for both groups. Both experts and novices were more accurate for a middle range of SFs as characterized by a quadratic function in which accuracy decreased toward the LSFs and HSFs, and all participants were most accurate with the full spectrum. However, the experts, but not the novices, showed a similar quadratic relationship in RTs in which RTs increased toward the LSFs and HSFs. The RT distribution analysis showed the experts utilized information in the midrange SFs for family-level recognition irrespective of how quickly they responded (i.e., the quadratic trend was observed in all quartile RT bins). More importantly, responses on none of the SF conditions were faster than responses on the All SF condition, which contains the entire SF spectrum. Thus, bird novices utilized midrange SFs to make accurate responses but there was no systematic pattern in which SFs influenced how quickly they responded. Unlike novices, bird experts systematically utilized the middle SF range for making fast and accurate family-level recognition.

Experiment 2

The default mode of recognition of experts is at the most specific category level (Johnson & Mervis, 1997). Thus, in Experiment 2 the experts were tested to investigate which SF range they used at the *species* level, for example, MacGillivray's Warbler or Wilson's Warbler. Birds at the species level share, to a larger degree, external object shape relative to family-level birds, potentially increasing the role that internal information may play in recognition. As in Experiment 1, the participants were tested in a category verification task in which they judged whether a species-level category label matched a bird image or not. Based on the results of Experiment 1, we predicted that a midrange of SFs would be particularly important for recognition, and that this advantage would be present across the entire RT distribution.

Method

Participants. Fifteen expert birdwatchers, 20–63 years of age ($M = 41.00$, $SD = 14.61$), took part in Experiment 2. Nine of the 15 expert participants participated in Experiment 1, eight participated in a previous study examining effects of color on bird recognition (Hagen et al., 2014), seven participated in both, and four participated in neither of the previous experiments. The participants received monetary compensation for their participation. The data from one additional expert participant was lost because of technical issues.

Stimuli. The stimuli were selected from the Warbler (e.g., Yellow Warbler, Wilson's Warbler, and MacGillivray's Warbler) and Finches (e.g., House Finch, Pine Siskin, and American Gold finch) bird families (see Figure 5). Six species from each family and three exemplars of each species were selected. Thus, a total of 36 bird images were used in Experiment 2 (2 families × 6 species × 3 exemplars). The stimuli were collected from the Wahlheim, Teune, and Jacoby (2011) bird dataset and supplemented by images from the Internet that were independently verified by a bird expert.

Following the procedures used in Experiment 1, the bird images were transformed to create different SF conditions. Images were cropped and scaled to fit within a frame of approximately 250 × 250 pixels and pasted on a gray background using Adobe Photoshop CS4. Images subtended a visual angle of approximately 5.02 degrees vertically and 5.78 degrees horizontally.

Procedure. The experimental procedure was identical to Experiment 1, with the exception that six species of birds from the Warbler and Finch families were tested in each of the six SF conditions (2–4, 4–8, 8–16, 16–32, 32–64, and All cpi). Each experimental trial was repeated six times for a total of 432 experimental trials (2 families × 6 species × 3 exemplars × 2 YES/NO conditions × 6 repetitions). The trials were divided into three blocks of 144 trials and participants were provided with a rest break between blocks. The blocks were counterbalanced across participants. For YES trials, the species label (e.g., "MacGillivray's Warbler," "Wilson's Warbler") matched the subsequently presented picture. For the NO trials where the species label did not match the picture, the foil picture was selected from the same family as the species label (e.g., the label "Wilson's Warbler" was followed by a picture of a "MacGillivray's Warbler").
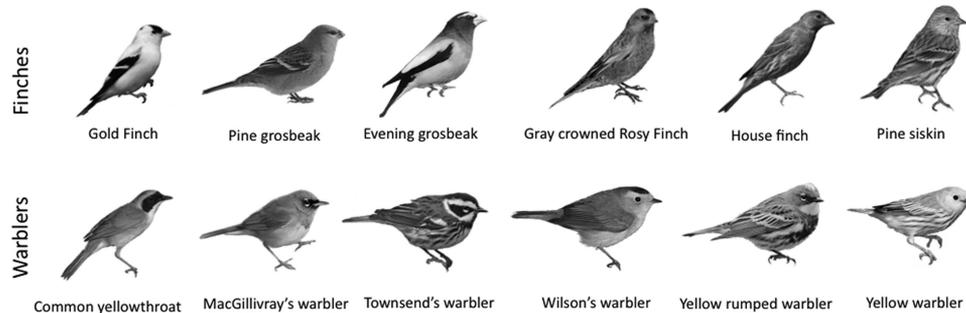

Figure 5. Examples of the stimuli used in Experiment 2. Note that these birds have a more similar global form than the birds used in Experiment 1.

Results

Similar to Experiment 1, trials with RT 3 *SD* greater than each participant's grand mean were excluded from further analyses. Excluded trials based on this requirement amounted to 75 out of 3,222 trials. No trials were excluded because of low accuracy with a given bird species (i.e., less than 50% accuracy). Following the same logic described in Experiment 1, our analyses were only performed on YES trials. See online supplementary material for analysis of the NO trials.

Accuracy. Figure 6 presents the mean accuracy data for YES trials as a function of SF condition. The accuracy data were analyzed in a repeated measures ANOVA using SF condition as a within-subjects factor. Consistent with Experiment 1, the main effect of SF condition was significant, $F(5, 70) = 59.50$, $p < .001$, $\eta_p^2 = .81$. Trend analysis revealed a significant quadratic trend, $F(1, 14) = 83.65$, $p < .001$, $\eta_p^2 = .86$, indicating that the experts were more accurate when the images contained the midrange SFs. Post-hoc comparisons revealed that participants were more accurate in the All SF condition than the other conditions (all *ps* < .005), with the exception of 8–16 cpi ($p = .646$) and 16–32 cpi ($p = .087$).

Response time. Figure 7 presents the mean correct RT data for YES trials as a function of SF condition. The RT data for the correct trials were analyzed in a repeated measures ANOVA using SF condition as a within-subjects factor. The main effect of this variable was significant, $F(5, 70) = 21.26$, $p < .001$, $\eta_p^2 = .60$. Similar to Experiment 1, trend analysis revealed a significant quadratic trend, $F(1, 14) = 56.30$, $p < .001$, $\eta_p^2 = .80$, suggesting that experts utilized a middle range of SFs to facilitate species-level categorization (compare Figures 3 and 7). Post-hoc comparisons revealed that participants responded more quickly in the All SF condition relative to the other conditions (all *ps* < .05).

RT distribution analysis. Similar to Experiment 1, we analyzed the distribution of correct RTs to examine when the quadratic trend observed in the overall analysis emerged. Figure 8 presents the mean correct RT data for YES trials as a function of bin and SF condition. The RT data were analyzed in a repeated
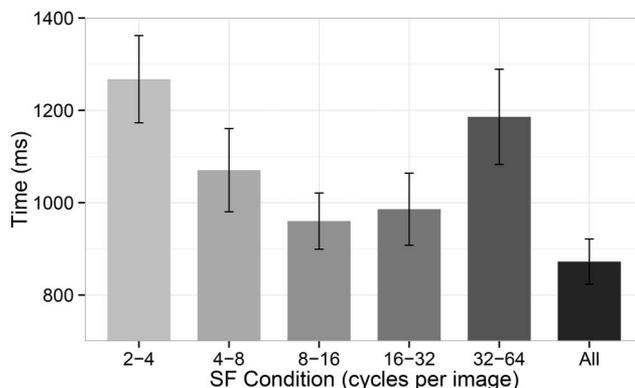

Figure 7. Experiment 2: Response time for the experts as a function of spatial frequency (SF) condition (2–4, 4–8, 8–16, 16–32, 32–64 cpi [cycles per image], and full spectrum). Error bars represent the *SEM*s.

measures ANOVA using SF condition and bin as within-subjects factors. The main effects of SF condition, $F(5, 70) = 21.17$, $p < .001$, $\eta_p^2 = .60$, and bin, $F(3, 42) = 58.81$, $p < .001$, $\eta_p^2 = .81$, were significant. The interaction between SF condition and bin was significant, $F(15, 210) = 8.59$, $p < .001$, $\eta_p^2 = .38$. There was a significant quadratic trend in each bin (bin 1: $F(1, 14) = 85.14$, $p < .001$, $\eta_p^2 = .86$; bin 2, $F(1, 14) = 99.16$, $p < .001$, $\eta_p^2 = .88$; bin 3: $F(1, 14) = 57.30$, $p < .001$, $\eta_p^2 = .80$; bin 4: $F(1, 14) = 27.53$, $p < .001$, $\eta_p^2 = .66$). Post-hoc comparisons revealed that in bin 1, participants responded more quickly in the All SF condition than the other conditions (all *ps* < .05) except 8–16 cpi ($p = .209$). In bin 2, they responded more quickly in the All SF condition than the other conditions (all *ps* < .01) except 16–32 cpi ($p = .051$). In bin 3, they responded more quickly in the All SF condition than the other conditions (all *ps* < .01) except 16–32 cpi ($p = .089$). In bin 4, they responded more quickly in the All SF condition than all the other conditions (all *ps* < .05).

The main finding of Experiment 2 was that SF filtering affected bird experts accuracy and RTs when they discriminated birds at the species level. As in Experiment 1, a quadratic trend was observed in which participants performed best for the midrange of SFs in term of both accuracy and RTs. Moreover, performance for the All SF condition, which contains the entire SF spectrum, was either as
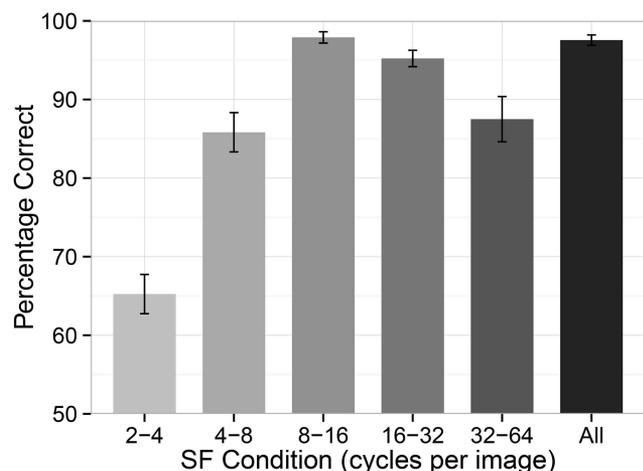

Figure 6. Experiment 2: Accuracy for the experts as a function of spatial frequency (SF) condition (2–4, 4–8, 8–16, 16–32, 32–64, and All cpi [cycles per image]). Error bars represent the *SEM*s.
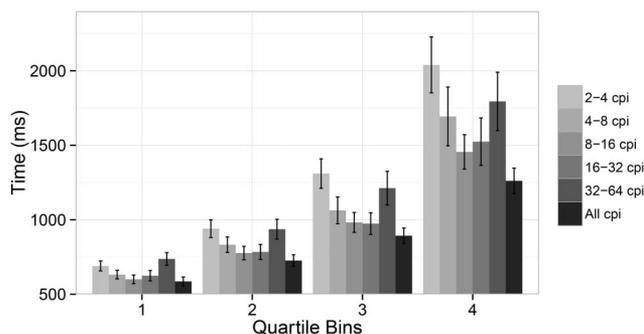

Figure 8. Experiment 2: Response time distribution for the experts. Bin 1 contains the 25% fastest responses of each participant. Bin 2 contains the next 25% fastest responses, and so on. Error bars represent the *SEM*s.

good as or better than any of the other filtered conditions. The quadratic trend in RTs was observed for all quartile RT bins. Thus, the internal information contained within a middle range facilitated species-level recognition relative to low and high SFs.

General Discussion

The aim of the current study was to test how *experience* with objects influences the role of internal object information for recognition. In Experiment 1, expert bird watchers and bird novices performed *family*-level recognition (e.g., Robin, Cardinal) of common birds that were band-pass filtered for different SF ranges and masked with the bird's contour. This combined manipulation preserved the birds' global form while systematically degrading its internal feature information. Consistent with previous work (Hagen et al., 2014; Johnson & Mervis, 1997; Tanaka & Taylor, 1991), bird experts were more accurate and faster at recognizing birds at the family level than novices (Figures 2 and 3). SF filtering affected the accuracy of the two groups similarly by facilitating recognition of images containing internal information that reside in a middle SF range as characterized by a quadratic function in which accuracy decreased toward the LSFs and HSFs. However, the experts, but not the novices, showed a similar quadratic relationship in RTs in which RTs increased toward the LSFs and HSFs. Thus, Experiment 1 demonstrated that bird experts, unlike novices, systematically utilized visual information contained in a middle SF range for fast yet accurate family-level categorization. In Experiment 2, only experts performed *species*-level categorization (e.g., Wilson's Warbler). Similar to Experiment 1, a midrange of SFs facilitated recognition in terms of both accuracy and RTs, across the entire range of RTs. Thus, similar to Experiment 1, a middle range facilitated species-level recognition.

Our results are consistent with previous studies exploring SFs in nonface object recognition (e.g., Collin, 2006; Collin & McMullen, 2005; Harel & Bentin, 2009). Collin and McMullen (2005) reported that subordinate recognition of everyday objects (including birds) relied more on internal features than basic-level recognition by showing that the former was disproportionately impaired by low-pass filtered images (i.e., only allowing SFs below 8 cpi). Similarly, the novices in our study were more accurate when the images contained internal features than external contour as indicated by higher accuracy in the midrange SFs than the LSFs. However, the novices' accuracy tapered off in the HSFs relative to the midrange SFs. Thus, by band-pass filtering the full SF spectrum into a number of narrow SF ranges, we extended Collin and McMullen's low- and high-pass approach by pinpointing the range that contained the optimal spatial resolution for internal details. The novice results suggest that the optimal range is defined by a midrange of SFs in which the internal details are not too spatially blurred nor too fine.

Similar to the novices, the experts were more accurate with the midrange of SFs relative to the LSFs and HSFs; however, unlike the novices, the experts were also faster to recognize the birds shown with the midrange SFs suggesting that they extracted from this range internal visual features (e.g., beak, wing-pattern, and eyes) more efficiently than novices. Interestingly, the experts used internal information contained in the midrange frequencies for both family-level (Experiment 1) and species-level (Experiment 2) recognition. These results indicate that internal information is

important regardless of whether the global form information provides good cues or relatively poor cues to recognition (compare the birds in Figure 1 and 5). Collectively, Experiments 1 and 2 suggest that a midrange of SFs provide the optimum spatial resolution for which internal object features are extracted, and importantly that *experience* can influence how recognition mechanisms operate within this range.

Together with previous work, our study suggests that experts quickly analyze several perceptual dimensions to improve recognition of homogenous objects within their domain of expertise. For example, Hagen et al. (2014) asked bird experts and novices to recognize birds at the family level to examine the role of color in expert recognition. Similar to the current study, the global form of the family-level birds were highly salient (see Figure 1) and could, therefore, be used for successful recognition. However, the experts still used color information for quick and accurate recognition. Similarly, Experiment 1 of the current study demonstrated that the experts used the internal features contained in the middle SF range despite the saliency of the bird's global form. This suggests that extensive experience discriminating homogenous objects produce subordinate object representation that contains both color and internal form information.

The results of our experiments are consistent with the idea that the features crucial for human object recognition reside in narrow ranges of SFs (Braje et al., 1995; Näsänen, 1999). Human object recognition relies on a small fraction (~1/10) of object information relative to ideal observers (Tjan et al., 1995). Moreover, computer simulations have shown that viewing the world through a band-pass SF channel allow for extraction of the features that enhance recognition (Braje et al., 1995). Consistent with this view, recognition of letters (Solomon & Pelli, 1994), geometrical shape objects (Tjan et al., 1995), and faces (e.g., Costen et al., 1994, 1996) rely on narrow bandwidths of SFs. In our study, the midrange SFs improved both novice and expert recognition emphasizing the importance of this narrow range of SFs for the internal features used for subordinate recognition. Our results are consistent with work showing that recognition is facilitated by narrow SF ranges; however, we also show that the *experience* of the observer can influence how efficiently those features are extracted and used to index objects in memory.

The results of our current study show interesting parallels and differences to the face domain—a domain in which most people have enough experience to be classified as experts (Tanaka, 2001). Previous work suggests that the information important for face recognition reside in a middle range (i.e., 8–16 cpi; e.g., Costen et al., 1994, 1996) that contains information about the relations among face features (i.e., configural information; Goffaux et al., 2005). This middle range is argued to be crucial for the holistic strategy that support face expertise (Maurer, Grand, & Mondloch, 2002). The midrange used by the experts in our study contained the 8–16 cpi. Thus, unlike Harel and Bentin (2013) who reported that face- and car-expertise rely on different SFs, our study indicate that different domains of expertise can rely on overlapping SF information. What is the role of these SFs in the 8–16 cpi in nonface expert object recognition? Similar to the role of these SFs in the face domain, it is possible that these SFs support a holistic strategy in experts. Indeed, nonface objects of expertise are sensitive to manipulations that impair configural processing, and so these SFs might support holistic recognition in the experts (Busey

& Vanderkolk, 2005; Campbell & Tanaka, 2014; Curby, Glazek, & Gauthier, 2009; Diamond & Carey, 1986; Gauthier, Curran, Curby, & Collins, 2003; Gauthier & Tarr, 2002; but see Robbins & McKone, 2007). However, our results also indicate that bird experts efficiently use coarse internal bird features represented by the 16–32 cpi SFs.

In summary, extensive experience in an object domain influence the way in which we encode and retrieve objects. Experiments 1 and 2 showed that a midrange of SFs provide the optimal resolution for extracting the crucial internal object information, and that as a result of extensive experience with birds, the information contained within the midrange was extracted and processed more efficiently. While object expertise is traditionally characterized by the ability to recognize objects of expertise equally fast at the basic and subordinate levels (Johnson & Mervis, 1997; Rosch et al., 1976; Tanaka & Taylor, 1991), our work shows that this downward shift in level of recognition is accompanied by the experts' ability to quickly use a larger repertoire of perceptual information, such as color (Hagen et al., 2014) and internal form information contained within the middle SF ranges (current study). Collectively, the studies suggest that experts recruit automatized *perceptual routines* to quickly analyze multiple perceptual dimensions. The perceptual routines are presumably facilitated by readily available object representations in which multiple perceptual dimensions (i.e., color, internal form) are encoded to increase the perceptual distinctiveness of objects in some high-level object space to improve subordinate-level recognition (e.g., Valentine, 1991).

References

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46. http://dx.doi.org/10.1037/h0027577

Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research, 35,* 2955–2966.

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the Inverse Efficiency Score (IES) a better dependent variable than the mean Reaction Time (RT) and the Percentage of Errors (PE)? *Psychologica Belgica, 51,* 5. http://dx.doi.org/10.5334/pb-51-1-5

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45,* 431–448. http://dx.doi.org/10.1016/j.visres.2004.08.021

Campbell, A., & Tanaka, J. (2014). Testing the face-specificity of the inversion effect in budgie experts. *Journal of Vision, 14,* 816. http://dx.doi.org/10.1167/14.10.816

Collin, C. A. (2006). Spatial-frequency thresholds for object categorisation at basic and subordinate levels. *Perception, 35,* 41–52. http://dx.doi.org/10.1068/p5445

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics, 67,* 354–364. http://dx.doi.org/10.3758/BF03206498

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception, 23,* 129–146. http://dx.doi.org/10.1068/p230129

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58,* 602–612. http://dx.doi.org/10.3758/BF03213093

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 94–107. http://dx.doi.org/10.1037/0096-1523.35.1.94

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117. http://dx.doi.org/10.1037/0096-3445.115.2.107

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195–201. http://dx.doi.org/10.1068/p120195

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6,* 428–432. http://dx.doi.org/10.1038/nn1029

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446. http://dx.doi.org/10.1037/0096-1523.28.2.431

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34,* 77–86. http://dx.doi.org/10.1068/p5370

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160. http://dx.doi.org/10.1111/j.0956-7976.2005.00796.x

Hagen, S., Vuong, Q. C., Scott, L. S., Curran, T., & Tanaka, J. W. (2014). The role of color in expert object recognition. *Journal of Vision, 14,* 9. http://dx.doi.org/10.1167/14.9.9

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1264–1273. http://dx.doi.org/10.1037/a0013621

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,* e67024. http://dx.doi.org/10.1371/journal.pone.0067024

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception, 15,* 595–602. http://dx.doi.org/10.1068/p150595

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience, 8,* 197–230. http://dx.doi.org/10.1162/jocn.1996.8.3.197

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211. http://dx.doi.org/10.3758/BF03212378

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126,* 248–277. http://dx.doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243–275. http://dx.doi.org/10.1016/0010-0285(84)90009-4

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience, 7,* 3416–3468.

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology, 2,* 165. http://dx.doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., Wong, A. C. N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2009). Time course of visual object categorization: Fastest does not necessarily mean first. *Vision Research, 49,* 1961–1968. http://dx.doi.org/10.1016/j.visres.2009.05.005

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260. http://dx.doi.org/10.1016/S1364-6613(02)01903-4

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8,* 454–469. http://dx.doi.org/10.3758/BF03196180

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39,* 3824–3833. http://dx.doi.org/10.1016/S0042-6989(99)00096-6

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103,* 34–79. http://dx.doi.org/10.1016/j.cognition.2006.02.008

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. http://dx.doi.org/10.1016/0010-0285(76)90013-X

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. *Journal of Vision, 1,* 413. http://dx.doi.org/10.1167/1.3.413

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369,* 395–397. http://dx.doi.org/10.1038/369395a0

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534–543. http://dx.doi.org/10.1037/0096-3445.130.3.534

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics, 61,* 1140–1153. http://dx.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? *Cognitive Psychology, 23,* 457–482. http://dx.doi.org/10.1016/0010-0285(91)90016-H

Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception & Psychophysics, 26,* 163–167. http://dx.doi.org/10.3758/BF03208310

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35,* 3053–3069.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology, 43,* 161–204. http://dx.doi.org/10.1080/14640749108400966

Wahlheim, C. N., Teune, R. K., & Jacoby, L. L. (2011). *Birds as natural concepts: A set of pictures from the Passeriformes order.* Retrieved from http://psych.wustl.edu/amcclab/AMCC%20Materials.htm