Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices

YETTA KWAILING WONG AND ISABEL GAUTHIER
Vanderbilt University, Nashville, Tennessee

Holistic processing (i.e., the tendency to process objects as wholes) is associated with face perception and also with expertise individuating novel objects. Surprisingly, recent work also reveals holistic effects in novice observers. It is unclear whether the same mechanisms support holistic effects in experts and in novices. In the present study, we measured holistic processing of music sequences using a selective attention task in participants who vary in music-reading expertise. We found that holistic effects were strategic in novices but were relatively automatic in experts. Correlational analyses revealed that individual holistic effects were predicted by both individual music-reading ability and neural responses for musical notation in the right fusiform face area (rFFA), but in opposite directions for experts and novices, suggesting that holistic effects in the two groups may be of different natures. To characterize expert perception, it is important not only to measure the tendency to process objects as wholes, but also to test whether this effect is dependent on task constraints.

Holistic processing—the tendency to process objects as wholes rather than as parts—is regarded as a hallmark of face recognition (Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Le Grand, & Mondloch, 2002; Young, Hellawell, & Hay, 1987). Holistic processing has different meanings in psychology, sometimes conceptualized as better recognition performance for object parts in the context of a whole object (the part–whole effect; Tanaka & Farah, 1993) or as evidence that patterns are not represented by a set of structural units (Chen, Allport, & Marshall, 1996). In the present study, we are concerned with a specific definition of holistic processing whereby observers are shown to be unable to selectively attend to part of an object (as in the composite effect; Young et al., 1987). Such failures of selective attention are associated with perceptual expertise for various nonface object categories, including cars (Gauthier, Curran, Curby, & Collins, 2003), fingerprints (Busey & Vanderkolk, 2005), and novel objects, such as Greebles and Ziggerins (Gauthier & Tarr, 2002; Gauthier, Williams, Tarr, & Tanaka, 1998; A. C.-N. Wong, Palmeri, & Gauthier, 2009). Holistic processing effects are also stronger for those faces with which we have the most experience, such as faces of one's own race (Michel, Rossion, Han, Chung, & Caldara, 2006; Tanaka, Kiefer, & Bukach, 2004) or of one's own age (de Heering & Rossion, 2008).

Although holistic processing is found to increase with perceptual experience, other work suggests that holistic effects can be found with unfamiliar objects. For example, holistic effects were obtained with novel objects (Greebles) when they were presented in the context of faces or when they were first encoded as two misaligned halves, rather than as a whole object (Richler, Bukach, & Gauthier, 2009). In another study, people with no expertise with Chinese characters processed them holistically, whereas experts did not (Hsiao & Cottrell, 2009). Holistic effects in novices appear to be inconsistent with an expertise account of holistic processing. However, are the causes of these holistic effects in novices similar to those in perceptual experts?

In the present study, we take a first empirical step toward addressing this question, since a more theoretical approach would be difficult without a consensus regarding the mechanisms underlying holistic processing, even in the case of face perception (Jacques & Rossion, 2009; Richler, Gauthier, Wenger, & Palmeri, 2008).[1] We defined *holistic processing* as a failure of selective attention, because this specific definition has been used both in studies in which holistic effects depend on expertise (e.g., A. C.-N. Wong, Palmeri, & Gauthier, 2009) and in others in which the effect was obtained in novices (Hsiao & Cottrell, 2009; Richler, Bukach, & Gauthier, 2009). We asked participants to selectively attend to part of an object (e.g., either the top or the bottom half) and ignore the irrelevant part. Holistic processing is thus indexed as the difference between congruent and incongruent trials, in which the irrelevant part led to either an identical or a conflicting response to that for the target part, respectively (the *congruency effect*, CE; Cheung, Richler, Palmeri, & Gauthier, 2008; Richler, Gauthier, et al., 2008; Richler, Tanaka, Brown, & Gauthier, 2008; A. C.-N. Wong, Palmeri, & Gauthier, 2009).

Y. K. Wong, yetta.wong@gmail.com

To compare holistic effects in novices and in experts, it is important to note that failures of selective attention occur in many different circumstances and tasks, but that not all of the CEs indicate a perceptual tendency to process objects as wholes (i.e., holistic processing). For example, one can obtain a CE when Greebles are viewed in the context of aligned faces but not for misaligned faces (Richler, Bukach, & Gauthier, 2009), suggesting that the CE for Greebles was induced under a specific context rather than reflecting a stable perceptual tendency. Also, with any selective attention paradigm, a CE may arise simply because of competing responses, similar to the classic Stroop effect, in which a reading response interferes with the response to ink color (MacLeod, 1991), which, again, is unrelated to perceptual tendency. In contrast, holistic effects obtained in perceptual experts—at least in the case of expert face perception—appear to be more automatic and remain relatively stable across various contexts. First, holistic effects for faces are relatively unaffected by task constraints that should influence the top-down deployment of attention—for instance, whether participants are asked to attend to one part of the face throughout the study (e.g., Michel et al., 2006) or to switch attention unpredictably to either the top or the bottom part of a face (e.g., Richler, Bukach, & Gauthier, 2009). Second, holistic effects for faces cannot be explained by Stroop-like response interference but can be explained by perceptual interference (Richler, Cheung, Wong, & Gauthier, 2009). Overall, this led us to hypothesize that holistic effects in experts are more automatic and more stable across various conditions, whereas holistic effects in novices can be obtained sometimes, but not at other times, depending on the specific contexts, strategies, and task constraints.

To test this hypothesis, we sought an object category for which both experts and novices might produce holistic effects (since in prior studies, novices or experts showed holistic effects on the same task but with different material). We chose to explore the domain of musical notation for a number of reasons. We thought that holistic processing may be engaged in expert music readers in the same way that it is for faces, because of the critical importance of spatial relationships between notes (Sloboda, 1978). Short sequences of musical notation are also roughly comparable in visual complexity to the Chinese characters that produced holistic effects in novices in Hsiao and Cottrell (2009), and, like these stimuli, these sequences are composed of parts that each hold limited meaning for novices. Fluency with musical notation varies greatly across individuals (even among experts), allowing us to examine how holistic effects vary as a function of music-reading skill.

Here, we tested the ability to selectively attend to part of a four-note sequence in music-reading experts and novices with a sequential matching task, and we tested whether it is stable across contexts prompting for different attention strategies. On each trial, two four-note sequences were presented sequentially, and participants were asked to judge whether the target note (one of the four notes cued by two arrows in the second sequence) was the same as or different from the equivalent note in the first sequence. As

in previous studies (Cheung et al., 2008; Richler, Gauthier, et al., 2008; Richler, Tanaka, et al., 2008; A. C.-N. Wong, Palmeri, & Gauthier, 2009), congruency was manipulated by shifting an irrelevant part (a note adjacent to the target note)—which led to a response congruent or incongruent with the correct response—and we used the CE to index the degree of holistic processing.

In two experiments, we created different contexts by manipulating target position (central or peripheral in the sequence) and target distribution (mostly at center, mostly in periphery, or evenly distributed in four positions) and, thus, encouraged participants to pay more attention to some notes than to others. This manipulation serves to reveal the extent to which the obtained holistic effects are dependent on various attention strategies. In particular, if the cause of holistic effects is strategic, the magnitude of the CE should be different as functions of target position and target distributions. Specifically, the CE should be small when targets appear in attended locations (encouraged by higher target distribution in those locations), because the distractors then fall into locations that are relatively ignored. When targets appear in locations that are relatively ignored, however, the distractors fall in areas that receive more attention and are, thus, more likely to interfere. Therefore, a contextual CE would be stronger when targets appear in relatively infrequent locations. In contrast, if holistic effects are relatively automatic, the magnitude of the CE should likewise be relatively similar across target positions and target distributions, even though different attentional strategies are expected. To preview our results, we found that holistic effects can be obtained with music sequences in both experts and novices. Across two experiments, we found that holistic effects were highly strategic for novices but were more stable across contexts in experts.

EXPERIMENT 1

Method

Participants

We recruited 66 participants (18–40 years of age) from Vanderbilt University and the Nashville community for course credits or cash payment. All participants reported amount of experience in music reading and rated their music-reading ability (1 = *do not read music at all*; 10 = *expert in music reading*). They were subsequently divided into an expert group (n = 25, 6 males, 19 females; years of experience ≥8 and a self-rating score ≥8), a novice group (n = 22, 7 males, 15 females; participants reported that they could not read music), and an intermediate group (n = 19, 8 males, 11 females; participants reported that they read music but did not meet the expert criteria). All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

Stimuli and Design

The experiment was conducted on Power Mac G4 computers using MATLAB (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Forty sets of four-note music sequences were generated with Finale Notepad (www.finalemusic.com/notepad/), and the notes were linked up with a straight line (Figure 1). The pitch of the notes was randomly generated, under the constraint that pairs of adjacent notes were at least two steps apart, so that changes would not alter the global contour of the music
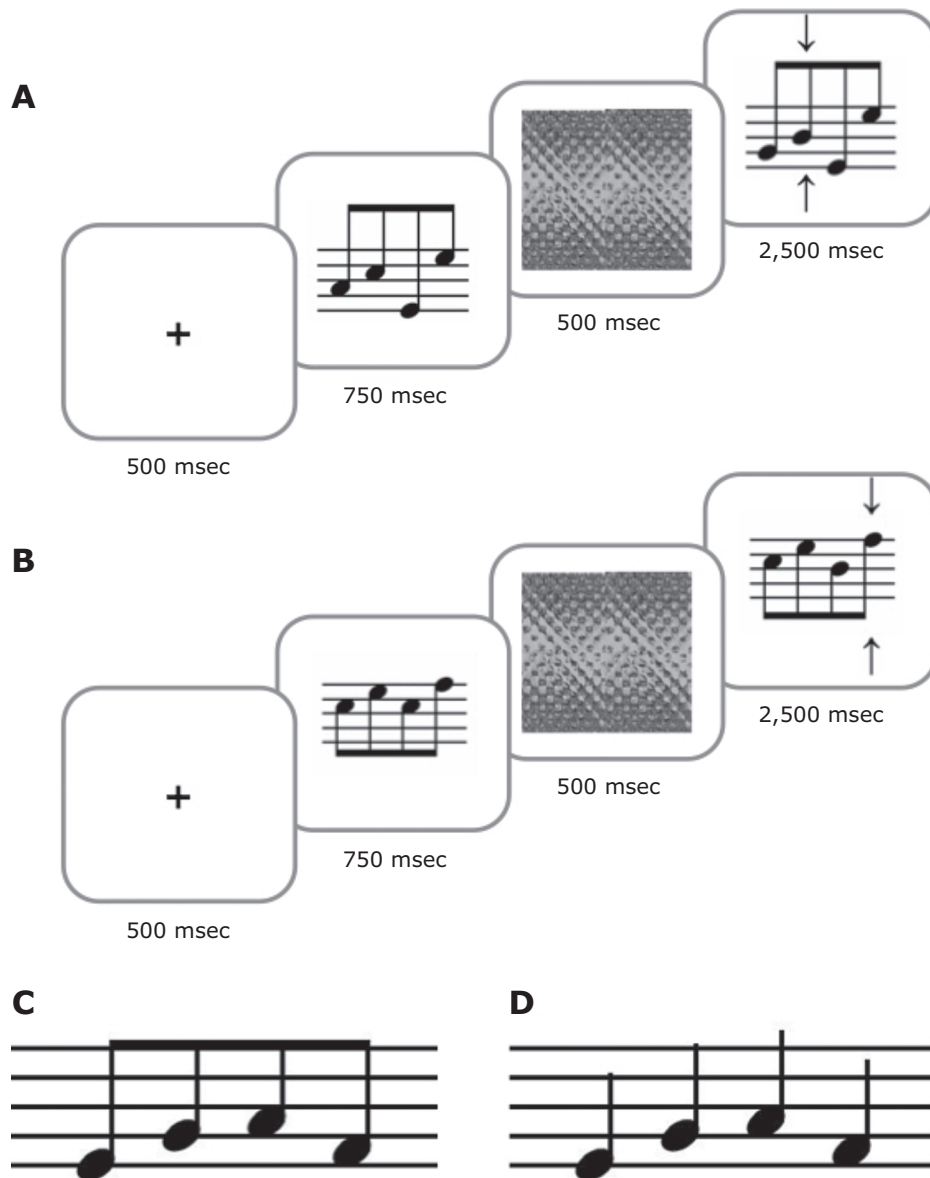
A

500 msec

750 msec

500 msec

2,500 msec

B

500 msec

750 msec

500 msec

2,500 msec

C

D

Figure 1. Sequential matching paradigm for Experiment 1. (A) A different–congruent trial (both the cued target note and an irrelevant note were shifted on the second sequence). (B) A same–incongruent trial (the cued target note did not change, but an irrelevant note was shifted on the second sequence). (C and D) Examples of the music sequences used in Experiment 2, in which the notes are either connected (panel C) or disconnected (panel D).

sequences. The notes ranged from the note below the bottom line to the note above the top line (corresponding to the range of D4 to G5 assuming a treble clef, while no clef was shown in the stimulus). All stimuli were black and were shown on a white background at approximately 4.8º × 4.8º of visual angle.

We used a sequential matching paradigm. On each trial, a fixation cross was shown for 500 msec, followed by a first note sequence for 750 msec, a mask for 500 msec, and a second note sequence for 2,500 msec (Figures 1A and 1B). A target note in the second sequence was cued with two arrows. Participants were asked to judge whether the target note was the same as or different from the equivalent note in the first sequence. Half of the trials were *same* trials, with the target note unchanged; the other half were *different* trials, with the target note shifted one step up or down. Participants were instructed to respond only according to the matching status of the target note by

a keypress. Both speed and accuracy were emphasized, and responses were required within 2,500 msec after the onset of the second sequence or were counted as errors (<1% of the trials).

Three factors were manipulated, with group as a between-subjects factor (experts, intermediates, novices) and congruency (congruent, incongruent) and target position (center, periphery) as two within-subjects factors. Targets appeared in the two center positions of the sequence (the second or third note) on 75% of the trials; they appeared in the peripheral positions (the first or fourth note) on 25% of the trials. This context was intended to encourage relatively more attention to notes in the central positions than to notes in the periphery. A note adjacent to the target was considered to be the distractor (left or right, counterbalanced, if the target was one of the central two notes). In the second sequence, the distractor note could be shifted one step up or down, resulting in different congruency conditions. Specifically, on
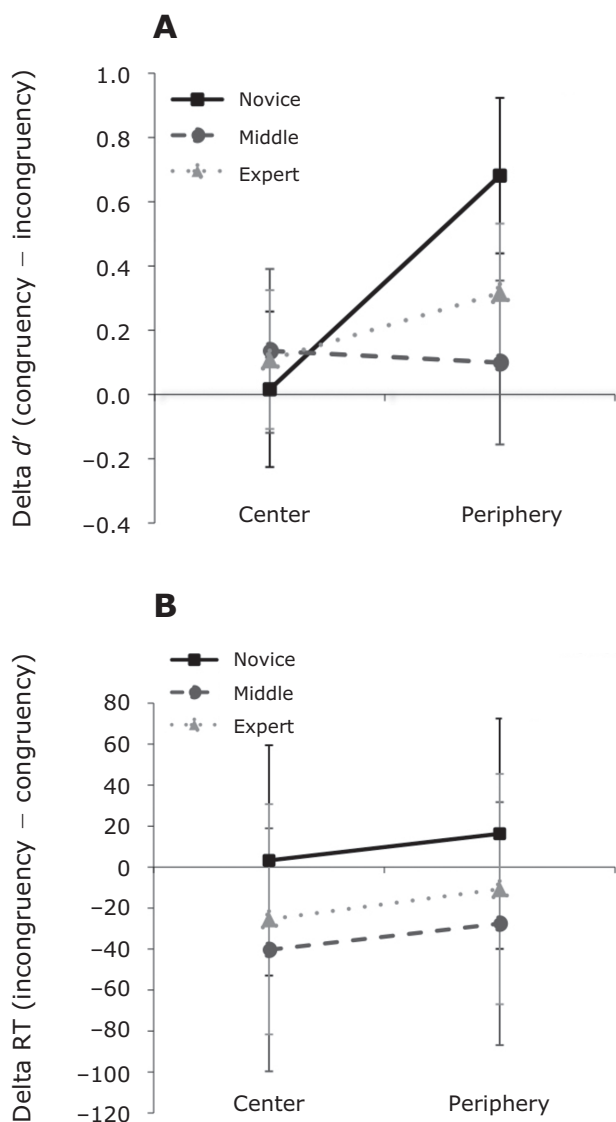
A

B

Figure 2. Mean delta *d′* (A) and delta response time (RT) (B) for Experiment 1. Error bars show the 95% CI for the within-subjects effects for the group × position interaction.

congruent trials, the distractor note remained unchanged (compared with the first sequence) on *same* trials, whereas it changed on *different* trials. For incongruent trials, the distractor note changed on *same* trials and remained unchanged on *different* trials. Dependent measures were sensitivity (*d′*) and response time (RT) for correct responses. Holistic processing was defined as the CE, using the difference in performance (*d′* or RT) between congruent trials and incongruent trials. If the CE for novices is due to context-dependent strategies, the CE should be modulated by target position, whereas, if it is relatively automatic, the CE for experts should be minimally modulated. There were a total of 256 trials, with 64 trials for each of the four within-subjects conditions. Ten practice trials with feedback were included, followed by test trials without feedback.

Results and Discussion

Data for 2 novices and for 1 intermediate music reader were excluded because their overall accuracy was lower than 60%. Examination of the stimuli postexperiment re-

vealed that shifting the position of the target or distractor note in the second sequence caused a change of the angle of the line (connecting the four notes) between the two sequences in 27 of the 256 trials. To eliminate the possible confound of the angle change with our manipulations, these trials were excluded from data analysis.

For each condition, the CE was computed using delta *d′* (congruent *d′* − incongruent *d′*) or delta RT (incongruent RT − congruent RT). A 3 × 2 ANOVA (group, position) on delta *d′* revealed a main effect of target position [$F(1,60) = 8.18$, $p = .006$], with a larger CE for periphery-target trials. The interaction between group and target position was significant [$F(2,60) = 4.17$, $p = .020$; Figure 2]. Scheffé tests ($p < .05$) revealed that the delta *d′* was larger for the periphery-target trials than for center-target trials only for novices. Also, the delta *d′* for the periphery-target trials was larger for novices than for the intermediate group ($p < .05$) and marginally larger than for the expert group ($p = .088$). No main effect or interaction reached significance for delta RT.

Planned *t* tests were performed for each condition to see whether the delta *d′* or delta RT was reliably different from zero—that is, whether the CE was significant in each condition. Results revealed that only the delta *d′*'s for novices and experts in periphery positions were significant ($p < .05$).

Our results suggest that the CE can be found with music sequences for participants with all levels of expertise. Critically, target position affected the CE for novices, but not that for experts, supporting our hypothesis that the CE was more strategic in novices than in experts.

EXPERIMENT 2

In Experiment 2, target distribution was manipulated parametrically to further test the hypothesis that the CE is more influenced by context in novices than in experts. Furthermore, we tested whether connecting the notes influences the CE by facilitating the spread of attention between notes within sequences (Egly, Driver, & Rafal, 1994). We also assessed whether the magnitude of the CE is predicted by music-reading ability (perceptual fluency with musical notation measured in both experts and novices) and years of experience in music reading (in experts only). Finally, we studied the neural correlates of the CE for music sequences by analyzing the results for 9 music-reading experts and for 9 novices who participated in the present study as well as in a previous fMRI experiment with musical notation (Y. K. Wong & Gauthier, 2010). We focused our analyses on the right fusiform face-selective area (rFFA), in which an increased activity is related to holistic processing of faces (Rotshtein, Geng, Driver, & Dolan, 2007; Schiltz & Rossion, 2006) and other objects of expertise (Gauthier & Tarr, 2002; A. C.-N. Wong, Palmeri, Rogers, Gore, & Gauthier, 2009).

Method

Participants

We recruited 19 participants who had more than 10 years of music-reading experience and/or considered themselves music-

reading experts as experts (8 males, 11 females; 18–32 years old) and 19 participants who reported being unable to read music as novices (4 males, 15 females; 18–33 years old) for cash payment. Similar to Experiment 1, experts had more years of experience in music reading (on average, 13.8 years for experts and 0.42 years for novices) and higher self-rating ability (on average, 8.84 for experts and 1.21 for novices on a 10-point scale). All reported normal or corrected-to-normal vision and gave informed consent, according to the guidelines of the institutional review board of Vanderbilt University.

Stimuli and Design

The experiment was conducted on Power Mac G4 computers using MATLAB (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). A total of 768 four-note music sequences were randomly generated using MATLAB, under the constraint that all the notes would have the same valid pointing direction before and after any shift. All stimuli were black on a white background and subtended a visual angle of about 7.2º × 4.8º.

A sequential matching paradigm was used, as in Experiment 1, except that the second sequence in each trial was presented for 2,000 msec instead of 2,500 msec. Five factors were manipulated, with group as a between-subjects factor (experts, novices) and with congruency (congruent, incongruent), target position (center, periphery), target distribution (25p75c, 50p50c, 75p25c; see below), and connection (connected, disconnected) as four within-subjects factors. The manipulation of congruency and target position was identical to that in Experiment 1. Target distribution had three levels and was manipulated across blocks of trials. Within each block, targets were distributed 25% on periphery and 75% at center (25p75c; the distribution used in Experiment 1), 50% on periphery and 50% at center (50p50c), or 75% on periphery and 25% at center (75p25c). The order of the three target distributions was counterbalanced across participants within each group. Participants were told about the target distribution immediately before each block. The four notes were either connected with a horizontal line parallel to the staff or not, for the connected or disconnected condition, respectively (Figures 1C and 1D). Participants were asked to ignore the irrelevant nontarget notes that might interfere with their judgment. Dependent measures were sensitivity (d') and RT for correct responses. There were a total of 768 trials, with 32 trials for each of the four within-subjects conditions. There were 10 practice trials with feedback before the test trials (without feedback).

Measure of Perceptual Fluency

We quantified perceptual fluency with music sequences in a sequential matching paradigm with four-note music sequences (Y. K. Wong & Gauthier, 2010). On each trial, a fixation cross was presented at the center of the screen for 200 msec, followed by a 500-msec premask, a target four-note sequence for a varied duration, and (after a 500-msec postmask) two four-note sequences appearing side by side, one identical to the first sequence and the other with one of the notes shifted by one step (randomly chosen out of the four notes, with the up/down shifts counterbalanced). The task was to select the matching sequence by a keypress. We estimated a perceptual threshold using QUEST (Psychtoolbox; Watson & Pelli, 1983) as the duration of the target sequence required to keep performance at 80% accuracy. Sequences were randomly generated, using notes ranging from the note below the bottom line to the note above the top line (corresponding to the range of D4 to D5 assuming a treble clef, while no clef was shown in the stimulus). Contrast for all the stimuli was lowered by about 60% to avoid a ceiling effect. The threshold was measured twice, each with 100 trials, and the two thresholds were averaged.

To control for individual differences not specifically tied to expertise with notes, perceptual fluency for 4-letter strings was also measured in an identical procedure. The 4-letter strings were randomly generated with 11 letters—b, d, f, g, h, j, k, p, q, t, and y—selected because they contain parts extending upward or downward

(i.e., ascendants and descendants, respectively), similar to musical notation. To create the distractor string, 1 of the 4 letters was chosen (counterbalanced across stimuli) and replaced by a different letter randomly drawn from the set. The string stimuli were also shown with the same lowered contrast as the note sequences.

Results and Discussion

One novice was excluded from further analysis because of an overall accuracy less than 60%. As in Experiment 1, the CE was measured with delta d' (congruent d' − incongruent d') or delta RT (incongruent RT − congruent RT). Examination of the data revealed no speed–accuracy trade-off.

A 2 × 3 × 2 × 2 ANOVA (group, distribution, target position, connection) was performed on both delta d' and delta RT. None of the main effects or interactions involving connection reached significance (all $ps > .05$), and therefore, we summarize here the results of 2 × 3 × 2 ANOVAs (group, distribution, target position) on delta d' and delta RT.

In terms of delta d', the main effect of group approached significance [$F(1,35) = 3.83$, $p = .058$], with the CE larger for novices than for experts. Also, a main effect of distribution was obtained [$F(2,70) = 4.49$, $p = .015$], which interacted with group [$F(2,70) = 3.57$, $p = .033$], with the CE for experts unaffected by target distribution, whereas that for novices was higher for 25p75c than for 75p25c (Figure 3A).

For delta RT, a main effect of distribution was significant [$F(2,70) = 4.57$, $p = .014$], which interacted with target position [$F(2,70) = 4.16$, $p = .020$]. More interestingly, the three-way interaction (group, distribution, target position) was significant for delta RT [$F(2,70) = 5.96$, $p = .004$; Figure 3B]. We subsequently performed the 3 × 2 ANOVA (distribution, target position) separately for the two groups. For novices, the interaction between distribution and target position was significant [$F(2,34) = 7.51$, $p = .002$; Figure 3B]. Scheffé tests ($p < .05$) revealed that the CE for novices was modulated by target likelihood, because the CE increased when the target appeared in relatively unexpected positions (for 75p25c, the CE was larger for center-target trials than for periphery-target trials; for 25p75c, the CE was larger for periphery-target trials than for center-target trials). Also, the CE for periphery-target trials was larger for 25p75c than for 50p50c and 75p25c. For experts, the CE was affected by distribution [$F(2,36) = 3.41$, $p = .04$; Figure 3B]. Post hoc tests (LSD, $p < .05$) revealed that the CE was larger for 25p75c than for 75p25c. No other main effects or interactions reached significance.

Planned t tests were performed for each condition to see whether the delta d' or delta RT was reliably different from zero—that is, whether the CE was significant in each condition. Results revealed that novices produced significant CEs ($p < .05$) in various conditions on both delta d' (center-target trials in 25p75c, periphery-target trials in 50p50c and 75p25c) and delta RT (center-target trials in all target distributions, periphery-target trials in 25p75c). For experts, the CE was significant in some conditions on delta RT (periphery-target trials in 25p75c, center-target trials in 25p75c and 50p50c; $p < .05$).
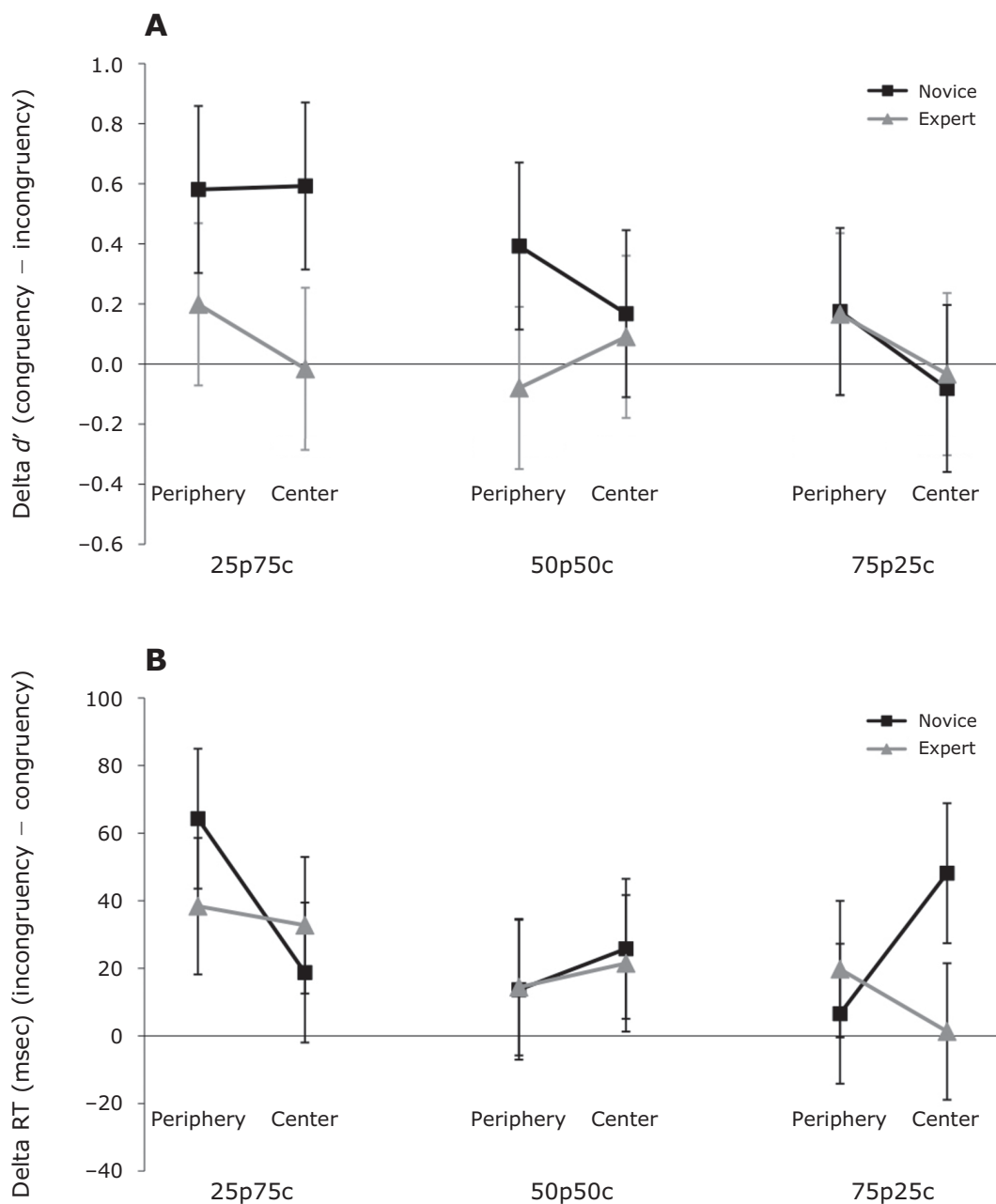
Figure 3. Mean delta d' (A) and delta response time (RT) (B) for Experiment 2. Error bars show the 95% CI for the within-subjects effects for the group × position interaction.

Our results replicated findings in Experiment 1, in that significant CEs were found for both experts and novices, with a stronger CE for novices than for experts. Both experiments converged, to suggest that the CE in novices was modulated by target position. However, the effect was obtained with delta d' in Experiment 1 and with delta RT in Experiment 2. It could arise from the differences in the design of the two experiments, such as (1) the target distribution was implicitly manipulated in Experiment 1, whereas participants in Experiment 2 were explicitly informed about the target distribution manipulations; or (2) there was only one target distribution manipulated in

Experiment 1, whereas, in Experiment 2, there were three (manipulated in different blocks), which might have led to adjustment of strategies for novices.

Both experts and novices were influenced by target distribution, with larger CEs in blocks in which targets were more frequent in the center than in the periphery of the sequence (this effect was found in delta d' in novices and delta RT in experts). This could be due to the fact that the distractor note was always adjacent to the target note. When periphery targets were more frequent, distractors never appeared in any of the attended periphery positions. However, in blocks in which central targets were more

frequent, distractors appeared in one of the more attended central positions for half of the trials, possibly leading to a larger CE for all participants.

More importantly, delta RT revealed that the CE for novices depended on whether the target appeared in an expected position (determined by both target distribution and target position), supporting our hypothesis that the novice CE is caused by context-dependent attentional strategies. In contrast, the absence of this target likelihood effect for experts is consistent with our hypothesis that the CE for experts is more automatic.

Slope Analysis

To quantify the degree to which holistic effects are dependent on context, we examined the extent to which the CE was modulated across different target distributions in each group. For each participant, the slope of delta d' (slopeDeltaD$'$) and delta RT (slopeDeltaRT) for either the center or periphery positions was calculated across the distributions of 25p75c, 50p50c, and 75p25c using the least-square method, with the linear weights of $[-1, 0, 1]$ for each distribution, respectively.

We performed a 2×2 ANOVA (group, target position) on slopeDeltaD$'$ and slopeDeltaRT separately. The main effect of group was significant for slopeDeltaD$'$ [$F(1,35) = 6.37, p = .016$], such that the magnitude of the slopeDeltaD$'$ was larger for novices than for experts (Figure 4A). For slopeDeltaRT, the main effect of target position was significant [$F(1,35) = 7.31, p = .01$], and the effect of target position interacted with group [$F(1,35) =$
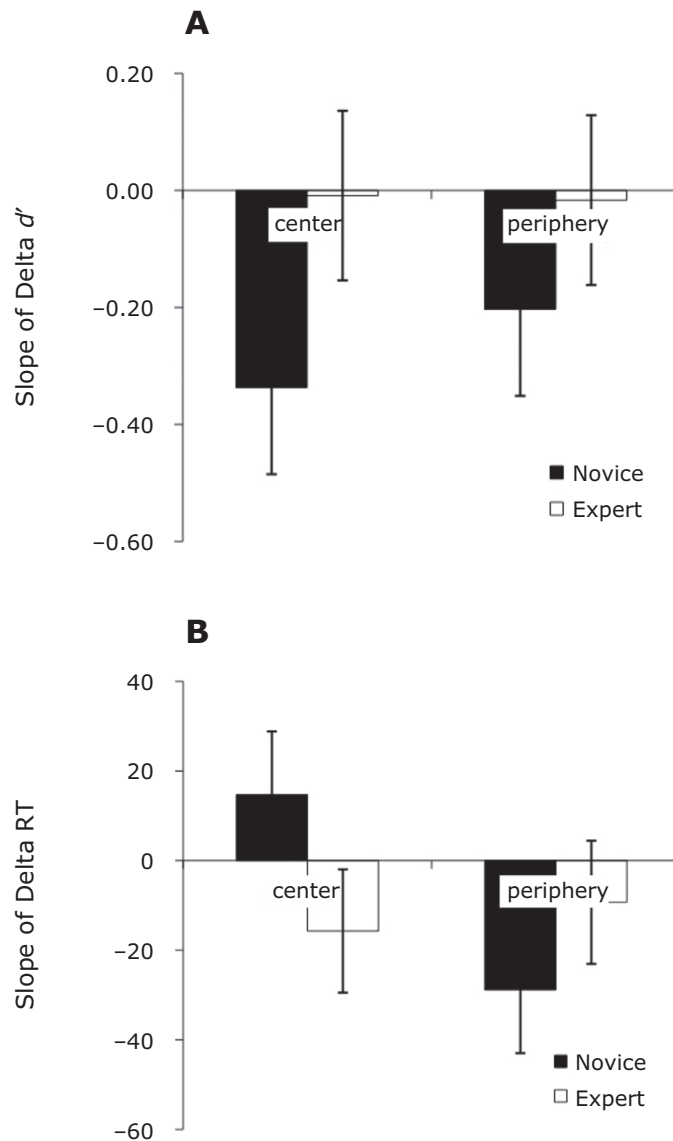

Figure 4. Slope of delta d' (A) and slope of delta response time (RT) (B) across the distributions of 25p75c, 50p50c, and 75p25c, generated by using the least-squares method. Error bars show the 95% CI within-subjects effects of the group × position interaction.

13.2, $p < .001$; Figure 4B]. Scheffé tests ($p < .05$) revealed that the slopeDeltaRT was larger for center-target trials than for periphery-target trials for novices, but it stayed similar across target positions for experts.

This further supports our hypothesis that the CE for novices is context dependent, whereas that for experts is more stable across different contexts.

Correlation Analyses

Conditions included in the correlation analyses. Our goal was to test the relationship between the CE and music-reading ability measured with perceptual fluency, years of music-reading experience, and the neural activity in the rFFA. In our design, there were 12 conditions from which we could extract the CE data for correlation analyses (2×3; target position, target distribution, on delta d' or delta RT). To keep the number of statistical comparisons to a minimum, we reduced the number of contrasts by focusing on five conditions that reflected the largest group differences in the CE (based on the behavioral findings above). First, we focused on the CE measured with delta RT, because it expressed the most important contextual differences between the two groups. To capture the effects for experts (modulated by target distribution irrespective of target positions), we included the three target distribution conditions with the CE for the two target positions averaged. To capture the effects for novices (modulated by target likelihood), we recoded our conditions into three levels of target likelihood (unlikely, equal, likely), in which *unlikely trials* refers to periphery-target trials for 25p75c and center-target trials for 75p25c, *likely trials* refers to center-target trials for 25p75c and periphery-target trials for 75p25c, and *equal trials* refers to 50p50c for both center- and periphery-target trials. Note that the condition of 50p50c is identical to the equal condition.

To summarize, we included the CE measured with delta RT in five different experimental conditions (25p75c, 50p50c or equal, 75p25c, unlikely, likely) in the correlation analyses between the CE and perceptual fluency with musical notation (in both experts and novices), years of experience in music reading (in experts only), and neural activity in the rFFA.

Predicting the CE with perceptual fluency for music sequences. Two novices failed to participate in this part of the experiment. One novice and 2 experts were excluded because they had a perceptual threshold more than 3 *SD*s away from the mean of the rest of their group. Therefore, 17 experts and 16 novices were included in the correlational analyses.

As we expected, experts could match music sequences faster than could novices ($M_{experts} = 239.3$ msec; $M_{novices} = 868.4$ msec), which was confirmed by a significant one-way ANOVA for group on the perceptual threshold for music sequences [$F(1,31) = 102.5, p \leq .0001$]. In contrast, the perceptual threshold for letter strings did not differ between groups ($p > .1$).

We considered the relationship between delta RT in each of five diagnostic conditions and the difference between the perceptual threshold for notes relative to that

for letters (N − L). For experts, N − L was negatively correlated with delta RT for 75p25c ($r = -.566, p = .018$; Figure 5A). For novices, in contrast, N − L was positively correlated with delta RT in the same condition (75p25c; $r = .699, p = .003$; Figure 5A) and in the target-likely condition ($r = .647, p = .007$). Thus, the CE decreased with perceptual fluency in novices but increased with perceptual fluency in experts.

Predicting the CE with years of music-reading experience. To test whether the CE pattern in experts relates to musical training, the CE was correlated with years of experience in music reading (ranging from 7 to 21 years) for the expert group only. Results revealed that more years of experience was associated with a larger CE with marginal significance ($r = .473, p = .055$), consistent with the possibility that the CE in experts could be due, at least in part, to musical training (rather than to a priori individual differences).

Predicting the CE with visual selectivity in the rFFA. To test whether holistic processing of music sequences is related to neural activity in the rFFA, as found for faces (Rotshtein et al., 2007; Schiltz & Rossion, 2006) and for other nonface object categories (Gauthier & Tarr, 2002; A. C.-N. Wong, Palmeri, & Gauthier, 2009), we analyzed the relationship between the CE in the present study and the neural selectivity for musical sequences investigated in an fMRI study, in which 9 experts and 9 novices from the present study also participated. Details of the fMRI methods can be found in Y. K. Wong and Gauthier (2010). Briefly, the neural selectivity for music sequences (five-note music sequences vs. five-letter strings and five-symbol strings) was measured within the group-defined rFFA (with separate localizer runs), while participants performed simple visual judgments on the stimuli (detecting either immediate repeats of images or a gap on the five-line background that existed in all stimuli). The results were combined across tasks, both in the original fMRI study and here, ensuring some degree of task-general selectivity in the fMRI measure.

For experts, delta RT was negatively correlated with neural selectivity for music sequences in the rFFA for 75p25c ($r = -.942, p < .001$; Figure 5B), and a similar negative correlation approached significance for two other conditions (50p50c, or equal trial, $r = -.648, p = .059$; unlikely trials, $r = -.605, p = .084$). For novices, in contrast, delta RT was positively correlated with neural selectivity for music sequences for unlikely trials ($r = .902, p = .001$; Figure 5C) and 25p75c ($r = .669, p = .049$). These results are surprising in many ways. Despite the fact that the rFFA was not selective for musical notation—even in individuals with musical experience (Y. K. Wong & Gauthier, 2010)—activity in this region predicts holistic effects with musical notation. In addition, this relationship goes in opposite directions for novices and experts.

GENERAL DISCUSSION

In two experiments, we demonstrated that holistic effects can be obtained—both in novices and in experts—

A

Perceptual Fluency for Experts and Novices

B   The rFFA for Experts

Neural Selectivity (Note – Letter)

C   The rFFA for Novices
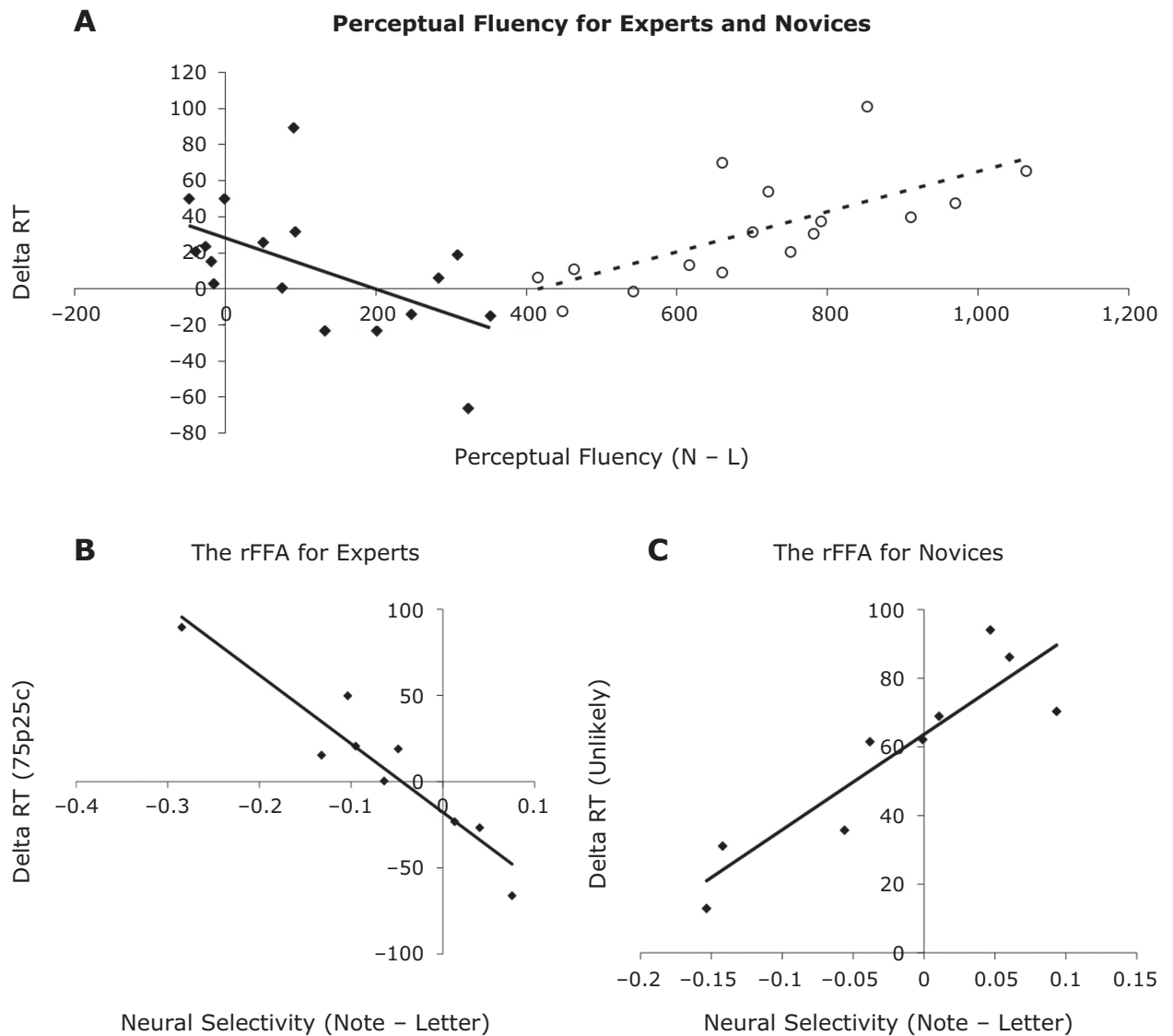

Neural Selectivity (Note – Letter)

Figure 5. Correlations between holistic processing and various behavioral measures and neural selectivity for notes in various visual areas, including perceptual fluency for experts (black dots and solid line) and novices (open circles and dotted line) in 75p25c (A), neural selectivity for music sequences in the right fusiform face area (rFFA) for experts in 75p25c (B), and that for novices in the *unlikely* condition (C). RT, response time.

with the same task and object category. Although the magnitude of holistic effects in at least some of our experimental conditions was predicted by music-reading ability and by neural activity for music sequences in the rFFA, the relationship ran in opposite directions for experts and novices. Our findings suggest that—even though holistic effects can reach a comparable (or even larger) magnitude in novices than in experts—the holistic effects for the two groups are of a different nature. Congruency effects in novices are highly dependent on experimental context, with effects that appear more strategic, whereas, in experts, they appear much more stable and automatic.

In novices, holistic effects may be caused by relatively slow processing of music sequences, forcing participants to attend to notes that correspond to the most probable target positions. This is suggested by the fact that the CE for

novices increased when targets appeared in unexpected positions and by the fact that perceptual fluency for music sequences is inversely related to the CE in novices (this strategy would be less necessary for novices who process notes more efficiently). In other words, and paradoxically, holistic effects for novices may be caused by higher selective attention to part of the objects, instead of by a failure of selective attention, as is the case in face perception (Richler, Tanaka, et al., 2008).

In music-reading experts, we observed relatively automatic holistic processing that increased with perceptual fluency for musical notes and with the number of years of musical training. This may reflect an increased tendency to process relative positions of notes in music sequences, as one becomes more proficient with musical notation. Because our congruency manipulations (i.e., the shift of

task-irrelevant notes) always change the relative positions of notes, these changes may be particularly salient to advanced music readers. Intuitively, music-reading experts may have developed increased sensitivity to the configuration of notes, to allow faster recognition of familiar patterns (e.g., scales or melodic lines of a piece), to enhance prediction of what different note sequences sound like (by processing the intervals between notes), and to enable effective planning of motor execution. In other words, music-reading experts may have developed the ability to identify both the individual parts and the configural information within a sequence, not unlike what is observed in face recognition (Hayward, Rhodes, & Schwaninger, 2008; Young et al., 1987).

It is important to realize that the present attempt to relate holistic effects to neural responses for notes was primarily exploratory—taking advantage of the same people participating in our behavioral and fMRI studies—and that it focused mainly on the ROI associated with holistic processing in prior work. Nonetheless, the correlations with activity in the rFFA are interesting for a number of reasons. First, they are consistent with correlations with perceptual fluency (and the rest of our behavioral results) in suggesting a difference in the nature of holistic effects for novices and experts. Second, they reveal that activity in the rFFA is not necessarily positively correlated with holistic processing. One possible explanation for the negative correlation of rFFA activity and holistic processing in experts is that, for expert readers of musical notation, the many codes (visual, auditory, motor) available for musical notation reduce reliance on this extrastriate area. After all, in our previous work, the FFA did not prove to be an area selective for musical notation in expert music readers (Y. K. Wong & Gauthier, 2010), consistent with the idea that different training experiences lead to different areas being recruited in the visual system (A. C.-N. Wong, Palmeri, Rogers, et al., 2009). In addition to revealing many areas recruited outside the visual system by expertise with notes, our prior fMRI work also uncovered expertise effects for notes in retinotopic cortex. Changes in early visual cortex could be related to the need for rapid processing of musical notation in both foveal and parafoveal visual regions or to other similarities between music-reading and perceptual learning tasks that recruit retinotopic cortex (Furmanski, Schluppeck, & Engel, 2004; Schiltz et al., 1999; Schwartz, Maquet, & Frith, 2002; Sigman et al., 2005). Further studies are needed to investigate the specific role of the FFA in holistic processing across a variety of domains.

Our findings address an interesting paradox in the study of holistic processing. Holistic effects have been found to increase with perceptual experience (Gauthier & Tarr, 2002; A. C.-N. Wong, Palmeri, & Gauthier, 2009) but have also been obtained in novices (Hsiao & Cottrell, 2009; Richler, Bukach, & Gauthier, 2009). We show that observing holistic *effects* is not sufficient evidence of underlying holistic processing, since holistic effects for notes in novices were shown to be context dependent and quite different from the automatic holistic processing that is considered the hallmark of object and face expertise. It

is possible that holistic effects in novices for Chinese characters (Hsiao & Cottrell, 2009) were likewise strategic, but this would need to be tested by varying different aspects of the task. In fact, we suggest that in investigations of holistic processing across various domains—including developmental studies or comparisons of different populations on face recognition skills or training studies with objects—it is important to examine both the magnitude of holistic processing and whether it varies across task and contextual manipulations.

REFERENCES

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

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

Chen, Y. P., Allport, D. A., & Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *Quarterly Journal of Experimental Psychology*, **49A**, 1024-1043.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception & Performance*, **34**, 1327-1336.

de Heering, A., & Rossion, B. (2008). Prolonged visual experience in adulthood modulates holistic face perception. *PLoS ONE*, **3**, e2317.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, **123**, 161-177.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, **105**, 482-498. doi:10.1037/0033-295X.105.3.482

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, **14**, 573-578.

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.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 431-446.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, **38**, 2401-2428.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, **106**, 1017-1027.

Hsiao, J. H., & Cottrell, G. W. (2009). Not all visual expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science*, **20**, 455-463. doi:10.1111/j.1467-9280.2009.02315.x

Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision*, **9**(6, Art. 8), 1-16. doi:10.1167/9.6.8

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, **6**, 255-260.

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, **17**, 608-615. doi:10.1111/j.1467-9280.2006.01752.x

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442. doi:10.1163/156856897X00366

Richler, J. J., Bukach, C. M., & Gauthier, I. (2009). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception, & Psychophysics*, **71**, 530-540.

Richler, J. J., Cheung, O. S., Wong, A. C.-N., & Gauthier, I. (2009). Does response interference contribute to face composite effects? *Psychonomic Bulletin & Review*, **16**, 258-263.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 328-342.

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 1356-1368.

Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience*, **19**, 1435-1452.

Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination. *NeuroImage*, **9**, 46-62.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage*, **32**, 1385-1394.

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences*, **99**, 17137-17142.

Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, **46**, 823-835.

Sloboda, J. A. (1978). Perception of contour in music reading. *Perception*, **7**, 323-331. doi:10.1068/p070323

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, **46A**, 2325-2345.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, **93**, B1-B9.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

Wong, A. C.-N., Palmeri, T., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a Ziggerin expert—but which type? *Psychological Science*, **20**, 1108-1117.

Wong, A. C.-N., Palmeri, T., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS ONE*, **4**, e8405.

Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, **22**, 695-713. doi:10.1162/jocn.2009.21229

Young, A. W., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception*, **16**, 747-759. doi:10.1068/p160747

NOTE

1. There is as of yet no consensus regarding whether holistic effects arise due to representational factors (e.g., parts not being explicitly represented) or to more decisional effects (e.g., dependencies in the decisions that use individually represented parts).