Ensemble Statistics Can Be Available before Individual Item Properties: Electroencephalography Evidence Using the Oddball Paradigm

Michael L. Epstein and Tatiana A. Emmanouil

Abstract

Behavioral studies have shown that statistical properties of object groups are perceived accurately with brief exposure durations. This finding motivated the hypothesis that ensemble perception occurs rapidly in vision. However, the precise timing of ensemble perception remains unclear. Here, we used the superior temporal resolution of electroencephalography to directly compare the timing of ensemble processing to that of individual object processing. The P3b was chosen as a particular component of interest, as it is thought to measure the latency of stimulus evaluation. Participants performed a simple “oddball” task in which sets of 51 lines with varied orientations sequentially flashed briefly on the display. In these sequences, there was a 20% chance of an individual oddball, wherein one marked object tilted clockwise, and a 20% chance of an ensemble oddball, wherein the average orientation of the set tilted 20% clockwise. In counterbalanced blocks, participants were instructed to respond to either individual or ensemble oddballs. ERP analysis was performed to test the timing of this processing. At parietal electrodes, P3b components were found for both individual and ensemble oddballs. Ensemble P3b components were found to occur significantly earlier than individual P3b components, as measured with both 50% area latency and 50% onset latency. Using multivariate pattern analysis, ensemble oddball trials were classifiable from standard trials significantly earlier in their time-course than individual oddball trials. Altogether, these results provide compelling evidence that ensemble perception occurs rapidly and that ensemble properties can be available earlier than individual object properties.

INTRODUCTION

In our daily life, we are frequently confronted with complex visual scenes containing many different groups of objects—crowds of faces, shelves of fruits, and vegetables. Although our ability to attend and memorize individual objects is limited (Luck & Vogel, 1997; Pylyshyn & Storm, 1988; Treisman & Gelade, 1980), we are able to quickly and easily extract general scene properties (Oliva & Torralba, 2006). How are the broader properties of the scene perceived despite the known limitations in encoding individual objects? A solution may be found in ensemble coding, a mechanism in vision by which groups of objects are summarized by their ensemble properties.

A breadth of evidence provides support for ensemble coding as a robust visual mechanism that operates independently of limited capacity resources over a variety of visual features. Previous research shows that ensemble perceptions are available even with brief stimulus presentation durations (Haberman & Whitney, 2007; Chong & Treisman, 2003) and without relying on attentional (Ji, Rossi, & Pourtois, 2018; Alvarez & Oliva, 2008) and working memory (Bauer, 2017; Epstein & Emmanouil, 2017) resources. Ensemble coding has also been shown to be possible for a variety of simple and complex visual features, such as orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), color (Maule, Witzel, & Franklin, 2014), size (Ariely, 2001), facial expression and gender (Haberman & Whitney, 2007), and even the overall “animacy” of a group (Yamanshi Leib, Kosovicheva, & Whitney, 2016). Together, this evidence suggests that ensemble coding is likely to occur at early stages of visual information processing and possibly before the visual system has processed individual objects in the scene.

Several models of visual perception now incorporate ensemble perception as an integral part of visual analysis. Reverse Hierarchy Theory, for example, postulates that ensemble processing occurs during a rapid feedforward sweep that provides a general gist of the scene (Hochstein, Pavlovskaya, Bonnehe, & Soroker, 2015, 2018). This preliminary analysis is then followed by feedback allocating limited resources to areas or objects requiring focused processing (Hochstein & Ahissar, 2002). Treisman (2006) posited that ensemble perception results from attention deployed to a broad as opposed to a narrow attentional window. According to this model, feature information is registered preattentively and in parallel within dedicated feature maps. However, the type of information extracted from feature maps depends on the deployment of attention. Focused attention on individual locations allows the binding
of object features in the attended locations, whereas attention distributed across a set of objects outputs a pooled summary of object features. Treisman suggests that ensemble representations are generated rapidly in vision, as the pooling of features could be potentially processed entirely in a feedforward sweep of processing.

Together, these findings and theories lead to a similar prediction: that ensemble statistics are calculated quickly in vision, potentially even completing more rapidly than individual object perception. Despite this strong prediction, however, few studies have directly tested this hypothesis. Here, we compared the timing of ensemble perception to that of individual object processing using the high temporal resolution of EEG. A well-established paradigm in EEG research is the “oddball” task, wherein stimuli are sequentially presented on the screen, and participants are instructed to watch for rare unique targets (Polich, 2007). Presentation of the “oddball” target stimulus can evoke a P3b component in recorded EEG signal, a large amplitude deflection typically starting around 300–500 ms post stimulus onset. The P3b component is thought to mark perceptual processing reaching a “completed” state, possibly indicating the availability of the perceptual representation to working memory or other higher-level processing (Polich, 2007). As a result, the latency of the P3b component is often used to infer the temporal dynamics of underlying processing (Polich, 2007; Magliero, Bashore, Coles, & Donchin, 1984; Kutas, McCarthy, & Donchin, 1977).

Here, we use the “oddball” paradigm to test the timing of ensemble processing and compare it to the timing of processing individual objects. Participants viewed groups of variably oriented lines and were instructed to attend to either their overall average orientation, or to focus on a single target within the group highlighted by a subtle cue. In the ensemble condition, participants responded when they detected the overall average orientation tilted clockwise, and in the individual condition, they indicated when the individual item tilted clockwise. Both the ensemble and the individual oddball targets appeared rarely, and we were thus able to elicit both ensemble and individual P3b components. We reasoned that if ensemble processing completes more rapidly than individual object processing, then P3b latencies would be faster for ensemble compared to the individual object oddballs. Importantly, the stimuli were generated using identical methods across both conditions; the only difference was the property to which participants were instructed to attend and accordingly the stimulus changes that constituted target oddballs in each condition.

Consistent with our hypothesis, we found that P3b components elicited for oddball targets in ensemble perception exhibited a significantly earlier latency as compared to those elicited for oddball targets in the individual condition. Using multivariate pattern analysis (MVPA), we also found that both ensemble and individual oddball trials could be classified above chance levels from standard trials and that ensemble trials were classifiable at significantly earlier time points. Together, this evidence provides compelling support for the speed with which ensemble properties can be extracted from groups of objects.

METHODS

Participants

Data from four pilot participants were processed for a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Set with an alpha of .05 and a power of .80, the power analysis indicated 32 participants would be required. Thirty-nine participants were originally recruited for this study. Seven of these participants were removed from the study (four for failing eligibility requirements, one because of an inability to reach a satisfactory impedance level in EEG, one for declining to continue the experiment after the first block, and one for reporting upon completion that they did not follow task instructions). This left 32 participants included in the results presented here (11 female, 2 left-handed, average age = 22.13 years). All included participants who had normal or corrected-to-normal vision as tested with a Snellen pocket eye chart (20/30 cutoff). All participants were self-reportedly free of neurological, psychological, or substance abuse disorders. All methods were approved by the CUNY Institutional Review Board.

Experimental Paradigm

The task consisted of a simple oddball paradigm where participants responded to rare targets in a stream of displays containing sets of 51 oriented lines arranged into a circular configuration. Participants performed two blocks: an ensemble block in which they were asked to attend to the overall average orientation of all the items and respond to oddballs in which the average orientation shifted clockwise, and an individual block in which they were asked to attend to the orientation of a single item within the group that was cued by a subtle semicircle, and to indicate when that item tilted clockwise. In both blocks, 60% of trials contained standard stimuli, 20% contained an ensemble oddball where the average orientation was shifted clockwise, and 20% of trials contained an individual oddball where only the individual item was tilted clockwise. Thus, ensemble and individual oddball targets never coincided on the same display and stimuli were generated using identical methods across blocks. The only difference between the ensemble and individual blocks was the property of the sets to which participants were instructed to attend and, as a result, the stimuli that constituted targets in each condition.

Each block was further split into 10 subblocks to allow for brief rest periods throughout the experiment. Within subblocks, the individually cued item maintained the same location across trials to avoid attentional shifts that
could introduce a difference between the ensemble and individual conditions. Individual target location was restricted to the center 16 items, excluding the four closest to fixation, and the location was randomly selected for each subblock, ensuring that participants attended to a variety of locations over the course of the experiment.

Stimuli

See Figure 1 for an example of the stimuli used in this experiment. The circular shape was defined as an $8 \times 8$ grid, with the three locations in each corner removed. Lines were $8^\circ$ in length and were placed into this grid each separated by $1.2^\circ$ of space as measured from the center of each line. The location of each line was additionally jittered up to $0.38^\circ$ on each trial ensuring that stimulus location varied across trials. A small fixation cross was positioned in the center of the screen during the task.

To discourage strategies where oddballs are detected by searching for extreme individual items, standard and oddball stimuli were generated to have different means but identical ranges. This was done by first generating a standard set with 51 items spaced evenly on a log scale progressing from 0.01 to 0.99 and then passing this array through a logit function ($\log(\frac{1}{y} - \frac{1}{2})$). The outputted curve was then shifted to have a minimum value of 0 and warped to have a range of 0–50 to represent the desired range of orientations. The individual target item’s orientation value of $20^\circ$ was added to this set, creating a standard stimulus with an overall mean orientation of $14.68^\circ$. Individual and ensemble oddball stimuli were then generated by adjusting these standard values. To generate ensemble oddball stimuli, the standard values were inverted (by subtracting each item’s value from 50 except for the individual target) ensuring that the distribution, range, and individual target matched, but with a new mean orientation of $35.13^\circ$. This resulted in a difference of $20.45^\circ$ between the standard and oddball averages. To generate individual oddball stimuli, only the individual target’s orientation was shifted a matching $20.45^\circ$ clockwise, and the remaining stimuli were correspondently shifted a small and evenly divided amount counterclockwise to maintain the same overall average orientation. These values were then randomly assigned to positions in the grid described above, with the individual target orientation always placed in the cued location. Because of a minor error in stimulus production discovered during analysis, on each trial, the far rightmost item in each set was not displayed, resulting in one random value from the calculated sets being excluded. Thus, although the description above accounts for 52 stimuli, only 51 were displayed on each presentation. Importantly, the missing item was always peripheral (far top-right of the set), and so the individual target could never be affected. Careful analysis of the remaining stimuli confirmed that the exclusion caused less than $1^\circ$ of jitter in the average orientation of the displays. Because of this, it is expected to have had an inconsequential effect on results in both conditions.

Procedure

After placement of the EEG cap, participants were seated comfortably in a dimly lit room. Stimuli were presented on a monitor set to a 60 Hz refresh rate. Participants completed the ensemble and individual blocks, with the order counterbalanced across participants. Participants were clearly instructed to keep their eyes focused on the fixation, and the on-line recording was monitored for eye movements to ensure that they followed these instructions. Each block contained 10 subblocks of 75 trials. On each trial, the display of oriented lines was displayed for 200 msec followed by a 1250- to 1750-msec interstimulus interval with only the fixation cross present. Participants responded to individual and ensemble oddballs by pressing the right arrow key. Between subblocks, participants were given a break screen to prevent fatigue and eye strain. When they were ready to continue, they could press any key to resume the experiment. Overall, for each condition (ensemble, individual), participants completed 750 trials, with 150 being oddball and 600 being standard displays. Each condition was preceded by a practice session (the length of one or two subblocks) to ensure participants understood and were prepared for the task.

EEG Recording and Preprocessing

EEG data were collected using a Neuroscan 64 channel Quick-cap, with electrodes organized using the extended 10–20 system using CZ as an on-line reference. Impedance for all electrodes was verified to be below 25 kΩ before beginning recording for each condition. Data were sampled on-line at 1000 Hz. Off-line analysis was carried out using the FieldTrip and EEGLAB toolboxes and custom written MATLAB scripts (Oostenveld, Fries, Maris, & Schoffelen, 2011; Delorme & Makeig, 2004). Only trials with correct responses were used. Trials with missed oddballs, or false alarm responses to standards, were excluded for the EEG analysis. Epochs were defined as 500 msec before to 1000 msec after stimulus onset. Data were filtered from 0.1 to 100 Hz using two-pass Butterworth filters, as well as...
notch filtered to remove 60-Hz line noise. Data were referenced to linked mastoids, linearly detrended and baseline corrected to the prestimulus activity. Data were next subjected to independent component analysis using EEGLAB’s ‘runica’ function, allowing us to visually identify and remove artifacts related to eye-blinks and temporal muscle movements for each participant. A final step of artifact rejection was carried out by calculating the absolute amplitude in each trial and excluding all trials exceeding ±80 mv. The full 500-msec baseline was used while cleaning data, as longer epochs have been shown to ensure the reliability of independent component analysis (Groppe, Makeig, & Kutas, 2009). After cleaning, data were re-baseline corrected to the 100 msec prestimulus period.

We planned to interpolate channels if overly noisy channels were found, but this was found to be unnecessary for any participants to reach a satisfactory level of cleaned data. We further planned to exclude participants who were left with fewer than 75 trials in any condition after cleaning, but all participants had sufficient data to surpass this threshold, and thus all participants were included in the study (average clean and correct trials used in analysis for ensemble standard = 397; for ensemble target = 127.5; for individual non-target in ensemble condition = 131.7; for individual standard = 391.6; for individual target = 125.28; for ensemble non-target in individual condition = 130.5).

**Behavioral Analysis**

In order to test for behavioral differences between conditions, we analyzed accuracy and RT. Accuracy was defined simply as the percentage of correct responses to an oddball out of the total number of oddball stimuli (150). Misses and false alarms were also recorded. RT was defined as the latency to response post stimulus onset. For RT analyses, response times with a z score ±2.5 were excluded. RT results in the behavioral analysis are reported using mean RTs. For the correlations between RTs and ERP latencies, we used the more appropriate median RTs, because of the different samples for each analysis, the grand averages displayed in all figures include all participants. To account for this, we excluded any participants whose peak measurements for a particular component exhibited the incorrect polarity (e.g., no positive activity in the defined P1 window). This left a still sufficient number of participants for analyses in all tests, with 23 participants for our P1 analysis, 30 for the N1, 22 for the P2, and 16 for the N2. All participants exhibited positive activity within the P3b window, and thus none needed exclusion for those analyses. Statistics reported here have the participants showing noisy data excluded; however, we found similar results when including all participants. Note that, because of the different samples for each analysis, the grand averages displayed in all figures include all participants.

**MVPA Analysis**

MVPA was implemented using the ADAM toolbox (Fahrenfort, van Driel, van Gaal, & Olivers, 2018) to compare classification accuracy for ensemble and individual oddball trials against their respective standard trials. The same preprocessed and cleaned data from our ERP analysis was used again for the MVPA, albeit resampled to 256 Hz to facilitate processing. No additional preprocessing steps, such as normalization, were added. Standard linear discriminant analysis was performed for both of these comparisons at each time point using a 10-fold cross-validation procedure using data from all scalp channels. In this procedure, data for each comparison were split into 10 equal folds, and classification was performed by training on 9 and testing on the remaining fold. Each individual fold was used once for testing. Performance was averaged across each of these tests. Because of the different number of trials in the standard
and oddball conditions, we performed oversampling using the ADASYN method (Fahrenfort et al., 2018). ADASYN corrects for class imbalances by generating synthetic data for the minority class, using a weighted algorithm ensuring that synthetic trials are close to the decision boundary. This serves to reduce bias by both balancing the size of the groups as well as shifting the classifier decision boundary toward difficult examples, while avoiding any possibility of repeating trials in training and testing folds (for a detailed description of these methods, see Fahrenfort et al., 2018; He, Bai, Garcia, & Li, 2008). Classification performance was measured using area under the curve (AUC), a more sensitive measure of classification accuracy than raw classification accuracy (Fahrenfort et al., 2018). AUC is taken from the receiver operating characteristic curve plotting true-positive against false-positive rates for each classification. Performance was then tested against chance (50%) to judge the time points at which the classifier was successfully able to differentiate oddball trials from standards. Cluster-based permutation correction with 1000 iterations was used to correct for false positives (see Fahrenfort et al., 2018; Maris & Oostenveld, 2007, for more details on this method). Finally, paired t tests on 50% onset latency measures were used to compare the latencies for successful classification of ensemble and individual trials.

In addition to the standard classification, we performed a temporal generalization analysis (King & Dehaene, 2014). Temporal generalization offers a way to visualize the dynamics of activity present in the signal by testing how well training for each individual time point extends to successful classification for the other time points within the signal. This can then be visualized using a color map plot, with the y axis representing training times and the x axis representing testing times. Here, the diagonal will show the typical classification (training and testing on the same time points) and the results off diagonal will show how well training extends to other time points. Temporal generalization used the same methods as described for the above analysis, but with testing repeated for each time point and 1000 iterations of cluster-based permutation testing to correct for false positives. To compare generalization across conditions, the output data for the individual condition was then subtracted from the ensemble condition (and once more corrected using cluster-based permutation testing) to create a difference map.

RESULTS

Behavioral Results

See Figure 2 for behavioral results. Analysis of the behavioral data revealed that participants were equally accurate at responding to ensemble and individual targets, indicating that difficulty was appropriately matched across conditions, \( t(31) = 1.43, p = .16, d = .25 \). RT was different between conditions, with responses to the ensemble oddballs being significantly faster despite participants being instructed to focus on accuracy rather than response speed, \( t(31) = 2.98, p = .0055, d = .53 \). No differences were found between conditions for false alarms, \( t(31) = .54, p = .60, d = .10 \), or for misses, \( t(31) = 1.43, p = .16, d = .25 \).

Together, these results support that the difficulty of each task was roughly similar, suggesting that any effects found on the latency and amplitude of the P3b component are unlikely to be due exclusively to difficulty in the task (for a more extensive discussion of difficulty effects, please see the general discussion). Furthermore, they provide initial support for the hypothesis that, despite difficulty being matched across conditions, ensemble properties may become available for response more rapidly than individual object properties.

ERP Results

See Figure 3 and Table 1 for ERP results. P3b components were successfully elicited by oddballs in both the ensemble and individual conditions. Critically, ensemble P3bs showed faster latencies as measured with 50% area (26.53-msec difference) and 50% amplitude onset measures (48.19-msec difference), as well as higher average amplitude (.55-mV difference). Similar effects were found for peak amplitude (1.27-mV difference) and latency measurements (42.31-msec difference). Measuring these effects in the P3b difference waves (oddball minus standard conditions), significant effects were again found for 50% area (27.5-msec difference) and for 50% onset amplitude (49.69-msec difference) latency, with peak latency trending. Importantly, differences in mean amplitude were not present for the P3b difference waves for either average or peak measurements. The earlier latency of the P3b component for ensemble compared to individual oddball targets is consistent with our hypothesis that
ensemble perception completes more rapidly than individual object perception. These results are unlikely to be explained by difficulty differences since the accuracy of the two conditions was found to be well matched.

No clear effects of the nontarget conditions (i.e., ensemble oddballs in the individual condition and individual oddballs in the ensemble condition) were found. First, these nontarget oddballs did not show a clear P3 component (see Figure 3A). Second, an ANOVA with oddball type (ensemble, individual) and task relevance (target, nontarget) as factors on 50% area latency showed no main effect of oddball type, $F(1, 31) = 1.15, p = .29, \eta^2_p = .036$, or task relevance, $F(1, 31) = 0.02, p = .90, \eta^2_p = .001$. An interaction between oddball type and task relevance was found, $F(1, 31) = 5.47, p = .026, \eta^2_p = .15$, explained by the finding that ensemble oddballs produced faster responses than individual oddballs when these stimuli were targets, $F(1, 31) = 12.09, p = .002, \eta^2_p = .28$, but not when they were nontargets, $F(1, 31) = 1.87, p = .18, \eta^2_p = .057$. These results suggest that oddball effects were dependent on task rather than stimulus intensity.

P1, N1, P2, and N2 components were also analyzed over occipital electrodes, excluding participants without clear components, to investigate possible differences in early visual processing (Figure 4; see Table 1 for statistics). No components showed any latency differences between ensemble and individual oddballs for 50% area, 50% amplitude, or peak latency measurements. P1 and N1 also did not show amplitude difference for either average or peak measurements. However, for the P2 and N2 components, we found significantly lower peak amplitude and mean amplitude in the ensemble condition as compared to the individual condition. Again, similar results were found for all measurements with all participants included. These differences in P2 and N2 amplitude may reflect differences in the distribution of attention between the ensemble and individual conditions, affecting early processing in early visual areas.

To further test the nature of the amplitude difference detected in the P2 and N2, we subjected each of these measurements to a $2 \times 2$ repeated-measures ANOVA with condition (ensemble vs. individual) and trial type (oddball vs. standard) as levels. Here, we found both an effect of trial type, with oddballs lower than their respective standards (P2: $F(1, 21) = 46.998, p < .001, \eta^2_p = .691$; N2: $F(1, 15) = 40.961, p < .001, \eta^2_p = .732$), as well as an effect of condition, with ensembles showing lower amplitude as compared to individuals (P2: $F(1, 21) = 75.017, p < .001, \eta^2_p = .691$; N2: $F(1, 15) = 19.102, p = .001, \eta^2_p = .56$). Interestingly, an interaction was also apparent with ensemble oddballs showing a larger difference compared to ensemble standards relative to the difference between individual oddballs and their standards (P2: $F(1, 21) = 8.916, p = .007, \eta^2_p = .298$; N2: $F(1, 15) = 10.426, p = .006, \eta^2_p = .410$). These results corroborate differences in P2 and N2 amplitude between the ensemble and individual blocks, possibly because of differences in attentional distribution between the conditions. Furthermore, because of the observed interaction, they also suggest that the differences are unlikely to be due solely to the distribution of attention. Rather, the specific process of processing an ensemble oddball appears to change signal amplitude as early as the N2 and P2 components.

Finally, we performed tests to see if P3b effects correlated with the RT difference found in the behavioral responses. For this test, we compared the 50% area measurements for each participant with their median response times, as median RT is considered to be an appropriately analogous measurement of the RT distribution (Luck, 2014). A test of Pearson correlations found significant correlations between both ensemble RT, $r(32) = .51, p = .003$, and individual RT, $r(32) = .62, p < .001$, and their respective P3b latencies.

**MVPA Results**

MVPA analysis was carried out to investigate the time course by which ensemble and individual oddballs could be distinguished from their respective standards. As can be seen in Figure 5, both were classified accurately above chance, with ensemble oddballs classifiable starting from 102 msec, and individuals classifiable starting from 210 msec. Measures of 50% onset latency supported that the ensemble condition could be classified earlier in the time course using both jackknife, $t(31) = -3.33, p = .002$, and individual, $t(31) = -2.34, p = .026$, measures. Note that one factor that may
| | Channels: CP1, CPZ, CP2, P1, PZ, and P2 | 
|---|---|---|
| **Latency Measures** | **Amplitude Measures** |
| | t | p | d | Average amplitude | t | p | d |
| P3 | 50% area latency | $-3.48$ | .002 | $-61$ | ** | 2.42 | .021 | .43 | * |
| $n=32$ | 50% amplitude onset latency | $-3.06$ | .005 | $-54$ | ** | 2.24 | .033 | .40 | * |
| | Peak latency | $-2.63$ | .013 | $-47$ | * | 2.42 | .021 | .43 | * |
| P3 difference | 50% area latency | $-3.21$ | .003 | $-57$ | ** | 1.11 | .27 | .29 |
| $n=32$ | 50% amplitude onset latency | $-3.56$ | .001 | $-63$ | ** | 1.11 | .27 | .29 |
| | Peak latency | $-2.01$ | .053 | $-36$ | † | 2.42 | .021 | .43 | * |

| **Occpital Components** | Channels: PO7, PO5, POZ, PO4, PO6, PO8, O1, OZ, and O2 | 
|---|---|---|
| **Latency Measures** | **Amplitude Measures** |
| | t | p | d | Average amplitude | t | p | d |
| P1 | 50% area latency | $-1.27$ | .22 | $-26$ | 1.03 | .32 | .21 |
| $n=23$ | 50% amplitude onset latency | $-0.99$ | .33 | $-21$ | .75 | .46 | .16 |
| | Peak latency | $-0.81$ | .43 | $-17$ | | 2.42 | .021 | .43 | * |
| N1 | 50% area latency | $-0.06$ | .95 | $-0.1$ | $-.22$ | .83 | $-0.04$ |
| $n=30$ | 50% amplitude onset latency | $0.33$ | .75 | $-0.06$ | $.69$ | .40 | $-0.07$ |
| | Peak latency | $0.88$ | $-0.50$ | $-0.12$ | | 2.42 | .021 | .43 | * |
| P2 | 50% area latency | $-0.04$ | .97 | $-0.1$ | $-5.04$ | $<.001$ | $-1.07$ | *** |
| $n=22$ | 50% amplitude onset latency | $1.3$ | .21 | $-0.28$ | $-4.45$ | $<.001$ | $-0.95$ | *** |
| | Peak latency | $0.05$ | $-0.96$ | $-0.01$ | | 2.42 | .021 | .43 | * |
| N2 | 50% area latency | $-0.08$ | .94 | $-0.02$ | $-4.70$ | $<.001$ | $-1.17$ | *** |
| $n=16$ | 50% amplitude onset latency | $-0.29$ | .78 | $-0.07$ | $-4.15$ | $<.001$ | $-1.04$ | *** |
| | Peak latency | $0.30$ | $0.77$ | $0.08$ | | 2.42 | .021 | .43 | * |

All statistics listed here are comparisons between ensemble and individual oddball ERPs.

* < .05.

** < .01.

*** < .001.

† Trending.
have contributed to high accuracy classification between oddballs and standards in both conditions is that oddballs required a motor response. However, this was equated between conditions, and would therefore not explain why ensemble oddballs could be classified earlier.

Notably, ensemble nontargets (ensemble oddballs appearing in the individual condition) could be classified from standards significantly above chance as well. Individual nontargets, however, still could not be classified above chance. Thus, although differences in the signal were not strong enough to be visible in the univariate ERP results, the increased power of whole brain MVPA revealed subtle differences in the processing of nontarget ensemble features. This provides further evidence that statistical changes in stimuli are detected easily, or potentially automatically, even when they are not the focus of the task at hand. It is important to note, however, that classification of ensemble oddballs was lower for nontargets compared to targets, which likely could be explained by the fact that nontargets were task irrelevant and did not require a motor response. This difference additionally implies that oddball responses were not driven solely by the physical properties of the oddball stimuli.

Similarly to the analysis described above for ERPs, we tested if the individual onset latencies measured with MVPA correlated with median RT. Tests of Pearson correlations again showed significant correlation for both ensemble, \( r(32) = .64, p < .001 \), and individual measurements, \( r(32) = .37, p = .035 \). In addition, we confirmed that onset latencies measured with MVPA further correlated with ERP onset latencies for ensembles, \( r(32) = .63, p < .001 \), and trended for individuals, \( r(32) = .34, p = .060 \), verifying that all measurements are closely related as predicted, and thus can be linked to the speed of processing.

Temporal generalization was also calculated to compare the dynamics of ensemble and individual item processing across time, the results of which are displayed in Figure 6.

Figure 4. Grand average ERP results taken from occipital electrodes showing the early visual components. Note that the displayed grand average includes all participants despite exclusions made in the analysis.

Figure 5. Classification accuracy for ensemble and individual oddball trials versus their respective standard trials shows that the ensemble condition was classified significantly above chance (solid lines) earlier than the individual condition. Measurements of classification accuracy for nontarget trials additionally shows that while ensemble nontargets were classified above chance, individual nontargets could not be successfully decoded. Classification accuracy is measured using AUC. Dashed lines show points of 50% onset latency for jackknife measures. For the ensemble nontarget condition, onset latency is displayed with an open dot on the x axis, as this time point was outside the significant cluster. Shaded areas around bars represent the standard error across participants.

Figure 6. The temporal generalization results (A) show that training the classifier at a given point of time generalized across a range of time points for both the ensemble (left) and the individual conditions (right). However, generalization started earlier in the ensemble condition, as can be seen by the contrast between the two generalization plots (B).
As can be seen in Figure 6B, which shows the significant differences between the temporal generalization for each condition, ensemble classification showed both an earlier onset and a wider window of generalization. These results support that, whereas both ensemble and individual conditions show some sustained activity during the processing of perceptual information, the ensemble condition displays a stronger and more sustained pattern of generalization.

**DISCUSSION**

In the current study, we compared the timing by which the visual system computes ensemble and individual object properties, using P3b components measured with EEG. P3b components showed significantly earlier latencies when participants processed ensemble oddballs, as compared to individual oddballs, within the same displays. In addition, RTs to ensemble oddballs were faster than to individual oddballs and were significantly correlated with the latency of the P3b components. We further found, using MVPA, that trials in which participants attended to ensemble oddballs could be classified accurately at earlier time points in the signal. Finally, we found no clear latency differences in the early evoked peaks observed over occipital areas, although we did observe lower amplitudes for the P2 and N2 components in the ensemble condition. Altogether, these results support the hypothesis that ensemble perception occurs rapidly and that the ensemble percepts become available more rapidly than individual object percepts.

P3b latency was selected as a primary measure in our task as it is a well-studied component thought to provide a useful index of when perceptual processes complete (Polich, 2007). Specifically, the P3b is often considered to mark stimulus detection and evaluation since its latency is influenced by such varied factors such as the presence of noise and distractors in the display, but not the compatibility between the stimulus and the required response (Magliero et al., 1984; Kutas et al., 1977). Although debate still stands as to the extent to which the interpretation of P3b latencies can be fully extricated from response-related processing (Verleger, Jaśkowski, & Wascher, 2005; Verleger, 1997), the component’s latency is nevertheless thought to be influenced by the timing of perceptual processing before response. Thus, it offers a useful avenue to explore the timing of ensemble perception in vision, by providing a well-defined index of when stimulus processing has reached a point where a functional statistical representation is available.

There were no differences in accuracy across conditions in our behavioral results, which suggest that our results cannot be attributed to differences in difficulty. Nevertheless, faster RT and P3b latency was observed for the ensemble condition, suggesting that even when difficulty was matched, ensemble processing could proceed more quickly. In other words, despite attending to all the items in the display and calculating an average property requiring information from all items, the overall average could be perceived more quickly than when focus was restricted to a single target. In addition, we found in our results a clear correlation between P3b latencies and RTs, providing further evidence that the faster perception of ensemble statistics can further be directly related to the ability to act more expeditiously in response to statistical information. Note that the faster RTs and P3b latencies for the ensemble condition could alternatively be interpreted as indicative of a difficulty difference. However, in the current study, we adopt the framework that difficulty is better measured by the discriminability of targets as indexed by accuracy and we allow that processing speed can vary between conditions of matched difficulty. Indeed, previous research has shown that, for visual tasks, the timing of behavioral and neural responses varies not only by difficulty but also based on the type of computations performed (Dobs, Isik, Panatzi, & Kanwisher, 2019; Lamme & Roelfsema, 2000; Navon, 1981). Therefore, given that our conditions were matched in accuracy, we attribute the differences found in RTs and ERP latencies to processing speed differences for ensemble and individual object computations. Nevertheless, it will be useful for future studies to examine the relationship between RT and accuracy measures, specifically in ensemble processing tasks.

Our analysis of visual components (P1, N1, P2, and N2) showed no differences in latency suggesting that the P3b differences are not a consequence of timing differences early in processing. This suggests that the timing of initial feature registration between the two conditions may be very similar. However, we did observe differences in amplitude between the ensemble and individual conditions, with ensemble oddballs producing lower amplitude. These differences are consistent with the possibility that ensemble and individual processing may result in different type of computations within early visual areas, for example, the pooling of activity in specifically feature-responsive areas of visual cortex (Hochstein et al., 2018; Treisman, 2006). Finally, we found an interaction between condition (ensemble, individual) and trial type (standard, oddball), such that the amplitude difference between standard and oddball trials was larger in the ensemble condition. This result suggests that visual cortex is involved in processing the ensemble properties of stimuli above and beyond the effects of attentional distribution.

The MVPA results further confirm and expand upon the latency differences observed in the ERPs. When trained on patterns drawn from the full scalp activity, ensemble oddballs could similarly be classified from standards significantly at earlier time points as compared to individual oddballs. The increased power of the MVPA revealed also that the ensemble oddballs were decodable as early as 106 msec, indicating that the differences in processing may begin earlier than observable in the ERP results. Interestingly the ensemble nontargets, which were indistinguishable from standards in the ERP analysis, could be accurately decoded...
with MVPA, supporting that statistical changes outside the focus of attention can be processed to some extent. Finally, temporal generalization provides a mapping of the dynamics of this process over time, indicating that the differences between ensemble and individual perception can be visualized as a sustained difference in processing. Together, these results provide additional nuance to the observations made in the ERP analysis.

It is necessary also to consider the differences in attentional distribution between our conditions and the role they could play in our results. Specifically, in the ensemble condition, participants distributed attention across all objects, whereas in the individual condition, they narrowly focused attention on the individual cued item. Ensemble coding and distributed attention have been considered to be largely intertwined (Jackson-Nielsen, Cohen, & Pitts, 2017; Treisman, 2006), with studies showing that ensemble processing benefits from attention distributed in space (Chong & Treisman, 2005) and that it can occur in some cases automatically (Corbett, Wurtsch, Schwartz, & Whitney, 2012) or with minimal attention (Alvarez & Oliva, 2008). Nevertheless, it is possible that attentional distribution affected activity independently of ensemble coding. Both visual N2 and P2 components have been well established to show modulation in their amplitudes related to the allocation of attention (Kanske, Ptitschka, & Kotz, 2011; Maeno, Gjini, Iramina, Eto, & Ueno, 2004; Johannes, Münte, Heinze, & Mangun, 1995; Van Voorhis & Hillyard, 1977). The amplitude differences of the P2 and N2 components observed here could indeed reflect differences in attentional distribution between the ensemble and individual conditions.

Any effects of attention are nevertheless unlikely to contradict the main results of our study. First, in comparing the parietal difference waves calculated between oddball and standard conditions (which were matched for attentional distribution), the amplitude difference between ensemble and individual conditions disappeared, but the latency effect remained. Furthermore, as mentioned above, the interaction observed in P2 and N2 components suggests that the presence of the ensemble oddball caused additional changes in their amplitudes, suggesting that attentional distribution cannot entirely be the cause of the differences found between tasks. Together, these results suggest that, although attentional deployment may have influenced overall amplitude differences between ensemble and individual conditions, it is unlikely to have accounted for the latency of P3b results or MVPA classification results, which relied on a comparison of standards and oddballs.

Note that the current study was designed with different oddball stimuli for each condition to ensure that participants attended to the required dimension and would not use ensemble and individual oddball information interchangeably across tasks. This manipulation introduces the possibility that stimulus intensity differences between ensemble and individual oddballs may play a role in our results regardless of task. However, analysis of nontarget oddballs showed that these conditions did not produce a P3 component and that they displayed significantly lower amplitude, suggesting that they elicited a qualitatively different response than the target oddballs. Furthermore, evaluation of response latency found that the differences between ensemble and individual oddballs were dependent on task and were eliminated when these same stimuli were task-irrelevant. Therefore, it is unlikely that our results were driven by stimulus differences between ensemble and individual oddballs.

The finding that ensemble perception can proceed more quickly than individual object processing has important ramifications for how ensemble perception is situated in theories of visual hierarchical processing. A number of theories position ensemble perception as an early type of processing—a first pass to help categorize stimuli (Utochkin, 2015) or otherwise provide a high level sense of the world (Hochstein et al., 2015; Hochstein & Ahissar, 2002). Our results lend support to a fundamental premise of these theories, showing that statistical outputs from visual processing become available more rapidly than individual object percepts. Strikingly here, the P3b latency was earlier even though the individual object property tested (orientation) was a relatively low-level feature. Although our results suggest that overall processing time for ensembles is faster than for individual objects, they do not pinpoint the exact stage at which this difference arises. However, these findings do provide some initial evidence that a difference may exist in early visual areas. Computational modeling of the expected activity in these areas, along with EEG and MRI paradigms, offers compelling future routes for future investigation.

Finally, it is worth addressing that, although these results provide support that ensemble estimates are available early, they do not exclude the possibility that ensemble perception estimates can be revised over time. There is, in fact, a growing body of evidence suggesting that ensemble perception dynamically changes over time. For example, multivariate analyses used in a previous EEG study provided evidence that ensemble representations for groups of faces are built progressively as information is collected (Roberts, Cant, & Nestor, 2019). Similarly, statistics can be accurately parsed across sequentially presented stimuli (Albrecht & Scholl, 2010; Haberman, Harp, & Whitney, 2009) and can be subject to serial dependence (Manassi, Liberman, Chaney, & Whitney, 2017). It has further been shown that the reported statistics often downweigh deviant items (Haberman & Whitney, 2010) and that these adjustments appear to be made via iterative steps of processing (Epstein, Quito-Dunn, Mandelbaum, & Emmanouil, 2020). It is possible that similar iterative computations take place for cases in which ensemble estimates are found to upweigh salient or target items within sets (Iakovlev & Utochkin, 2020; De Fockert & Marchant, 2008). The stimuli of the current study were not ideal for investigating the iterative nature of ensemble perception since they were
simple and performance was overall high. Indeed, temporal generalization in the MVPA analysis was found to be limited. Future studies using displays with outliers, multiple sets, or noisy representations are needed to more specifically test the iterative nature of ensemble perception using EEG.

Altogether, our study takes important steps in testing the timing of ensemble perception using neuroimaging. Despite a rich literature exploring the speed of ensemble perception using behavioral paradigms, few studies have, as of yet, tested the dynamics of this process directly with EEG. The present results complement and extend prior findings, supporting ensemble perception as a fast, early visual process that can even surpass the speed of individual item perception. These results offer compelling support for ensemble perception as an essential element of vision, providing a rapid summary of a frequently complex visual world.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Support for this project was also provided by a PSC-CUNY Award, jointly funded by The Professional Staff Congress and The City University of New York.

Reprint requests should be sent to Michael L. Epstein, Department of Psychology, City University of New York, 365 5th Ave, or via e-mail: mlepstein88@gmail.com.

Author Contributions

Michael L. Epstein: Conceptualization; Formal analysis; Investigation; Methodology; Writing—Original draft; Writing—Review & editing. Tatiana A. Emmanouil: Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Writing—Review & editing.

Funding Information

Michael L. Epstein, PSC-CUNY, grant number: TRADA-50-583; National Eye Institute (https://dx.doi.org/10.13039/100000053) grant number: R15EY030679.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (JoCN) during this period were M(an)/M = .408, W(oman)/M = .355, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article’s gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .609; W/M = .152; M/W = .196; W/W = .043.

REFERENCES

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world. Psychological Science, 21, 560–567. DOI: https://doi.org/10.1177/0956797610365545, PMID: 20424102

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. Psychological Science, 19, 392–398. DOI: https://doi.org/10.1111/j.1467-9280.2008.02098.x, PMID: 18399893, PMCID: PMC2587223

Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychological Science, 12, 157–162. DOI: https://doi.org/10.1111/1467-9280.00327, PMID: 11340926

Bauer, B. (2017). Perceptual averaging of line length: Effects of concurrent digit memory load. Attention, Perception, & Psychophysics, 79, 2510–2522. DOI: https://doi.org/10.3758/s13414-017-1388-4, PMID: 28779335

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. Vision Research, 43, 393–404. DOI: https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. Perception & Psychophysics, 67, 1–13. DOI: https://doi.org/10.3758/s13414-003-0009-0, PMID: 15912869

Corbitt, J. E., Wurmitch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. Visual Cognition, 20, 37–41. DOI: https://doi.org/10.1080/13506285.2012.657561, PMID: 24348083, PMCID: PMC3659532

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates independent component analysis. Journal of Neuroscience Methods, 174, 9–21. DOI: https://doi.org/10.1016/j.jneumeth.2008.09.006, PMID: 18102499

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. Nature Communications, 10, 1–10. DOI: https://doi.org/10.1038/s41467-019-09239-1, PMID: 30890707, PMCID: PMC6425020

Epstein, M. L., & Emmanouil, T. A. (2017). Ensemble coding remains accurate under object and spatial visual working memory load. Attention, Perception, & Psychophysics, 79, 2088–2097. DOI: https://doi.org/10.3758/s13414-017-1355-2, PMID: 28600677

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers. Journal of Experimental Psychology: Human Perception and Performance, 46, 1267–1297. DOI: https://doi.org/10.1037/xhp0000857, PMID: 32757952

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA using the Amsterdam decoding and modeling toolbox (ADAM). Frontiers in Neuroscience, 12, 368. DOI: https://doi.org/10.3389/fnins.2018.00368, PMID: 30018529, PMCID: PMC6087716

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175–191. DOI: https://doi.org/10.3758/BF03193146, PMID: 17695343
Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*, 411–443. DOI: https://doi.org/10.1080/13506280500195250, PMID: 17387378, PMCID: PMC1832144

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136. DOI: https://doi.org/10.1016/0010-0285(80)90005-5

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization primary categorization. *Journal of Vision, 15*, 1–14. DOI: https://doi.org/10.1167/15.4.8, PMID: 26317396

Van Voorhis, S., & Hillyard, S. A. (1977). Visual evoked potentials and selective attention to points in space.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology, 34*, 131–156. DOI: https://doi.org/10.1111/j.1469-8886.1997.tb02125.x, PMID: 9090263

Verleger, R., Jaskowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology, 19*, 165–181. DOI: https://doi.org/10.1027/0269-8803.19.3.165

Yamanshi Leib, A., Kosovichova, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications, 7*, 13186. DOI: https://doi.org/10.1038/ncomms13186, PMID: 27848949, PMCID: PMC5116093
AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please verify the changes made in the article title.
2. Please spell out CUNY.
3. Please spell out CZ if it is an acronym.
4. Please spell out both instances of ADASYN if it is an acronym.
5. Please define N.S.
6. Please verify the data in the GCBI section if correct.

END OF ALL QUERIES