Atten Percept Psychophys (2017) 79:100–116
DOI 10.3758/s13414-016-1216-2

Summary statistics in the attentional blink

Nicolas A. McNair[1] · Patrick T. Goodbourn[1,2] · Lauren T. Shone[1] · Irina M. Harris[1]

Published online: 13 October 2016

Abstract We used the attentional blink (AB) paradigm to investigate the processing stage at which extraction of summary statistics from visual stimuli ("ensemble coding") occurs. Experiment 1 examined whether ensemble coding requires attentional engagement with the items in the ensemble. Participants performed two sequential tasks on each trial: gender discrimination of a single face (T1) and estimating the average emotional expression of an ensemble of four faces (or of a single face, as a control condition) as T2. Ensemble coding was affected by the AB when the tasks were separated by a short temporal lag. In Experiment 2, the order of the tasks was reversed to test whether ensemble coding requires more working-memory resources, and therefore induces a larger AB, than estimating the expression of a single face. Each condition produced a similar magnitude AB in the subsequent gender-discrimination T2 task. Experiment 3 additionally investigated whether the previous results were due to participants adopting a subsampling strategy during the ensemble-coding task. Contrary to this explanation, we found different patterns of performance in the ensemble-coding condition and a condition in which participants were instructed to focus on only a single face within an ensemble. Taken together, these findings suggest that ensemble coding emerges automatically as a result of the deployment of attentional resources across the ensemble of stimuli, prior to information being consolidated in working memory.

✉ Nicolas A. McNair
nicolasmcnair@gmail.com

[1] School of Psychology, University of Sydney, Sydney, NSW 2006, Australia

[2] School of Psychological Sciences, University of Melbourne, Parkville 3010, Victoria, Australia

Keywords Attentional blink · Visual perception · Ensemble coding

Introduction

Despite the vast amount of incoming sensory information that reaches the brain, capacity limitations at various stages of processing place severe constraints on how much of this information is consciously accessed at any given point in time (Marois & Ivanoff, 2005). It has been suggested that the rapid summarisation, or "ensemble coding," of featural information shared between similar objects within a scene may provide a means by which the visual system mitigates some of the effects of these information-processing bottlenecks (Alvarez & Oliva, 2008; Chong & Treisman, 2003). Evidence of summary representations of stimulus features (almost universally, the average of such features) has been shown across a variety of stimuli, from simple features such as orientation (Oriet & Brand, 2013; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and size (Ariely, 2001; Chong & Treisman, 2003, 2005b) to high-level features like the emotional expression (Haberman & Whitney, 2007, 2009) and identity (de Fockert & Wolfenstein, 2009) of faces.

One of the key aspects of summary representations is that, despite their reasonable accuracy, recall of individual items within the ensemble is generally very poor, often at the level of chance (Ariely, 2001; Corbett & Oriet, 2011; Haberman & Whitney, 2009). This suggests that redundant featural information from the individual items might be quickly pooled into a summary representation, which is then retained, while the individual representations are lost (Chong & Treisman, 2003; Alvarez & Oliva, 2008). Such summary representations therefore could provide a means for circumventing the limited capacity of visual short-term memory (VSTM). Indeed, there

seems to be little or no cost associated with estimating averages from sets of stimuli in which the number of objects exceeds the 3- to 4-item limit of VSTM (Ariely, 2001; Attarha, Moore, & Vecera, 2014; Haberman & Whitney, 2009; Robitaille & Harris, 2011). Furthermore, such averages can be computed from displays presented as briefly as 50 to 100 ms (Chong & Treisman, 2003; Haberman & Whitney, 2009; Oriet & Brand, 2013), much faster than what would be expected if each item were serially encoded into working memory (Treisman, 2006). There also is some evidence that summary statistics are computed automatically before any information reaches awareness (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013, 2014), even for unattended stimuli (Alvarez & Oliva, 2008; Oriet & Brand, 2013). As such, the process has been likened to the low-level integration of visual information that provides the perception of texture (Im & Halberda, 2013).

Other evidence, however, positions ensemble coding at a later stage of visual processing. Chong and Treisman (2005a) found that ensemble coding was impaired when it followed a serial visual-search task requiring focused attention compared with a parallel visual-search task requiring distributed attention. In addition, they found that summaries were significantly less accurate when the spatial extent of attention required by a concurrent task did not include any of the items in the set, than when attention was spread over the entire ensemble. This dependence on the distribution of attention across the ensemble argues against a preattentive averaging mechanism. In support of this, pre-cueing the dimension to be reported produces equivalent benefits for single and ensemble stimuli, suggesting that the encoding of ensemble statistics requires similar attentional engagement as the encoding of single features (Huang, 2015).

Furthermore, the formation of summary representations of both size and orientation is affected by object-substitution masking (Jacoby, Kamke, & Mattingley, 2013). In this paradigm, processing of a target stimulus is impaired when it is surrounded by task-irrelevant dot stimuli that have an onset simultaneous with the target but a delayed, asynchronous offset (Di Lollo, Enns, & Rensink, 2000). Theoretical accounts of this effect suggest that it arises during late, re-entrant visual processing following the integration of featural information (Chakravarthi & Cavanagh, 2009; Di Lollo et al., 2000; Dux, Visser, Goodhew, & Lipp, 2010). Additionally, some findings that purport to show preattentive averaging may also be consistent with averaging occurring after the distribution of attention. For example, Alvarez and Oliva (2008) reported that participants were equally accurate in estimating the final centroid of both attended (target) and unattended (distractor) sets of moving dots. However, as the participants were aware that they would be asked to report the location of the distractor dots, they may have attempted to spread their attention across both sets. This is consistent with the finding that experimental manipulations resulting in the allocation of more attention to the target set had a negative impact on localising the centroid of the distractor set. In another study, Oriet and Brand (2013) found that unattended stimuli influenced the mean estimates of attended stimuli. However in that study, the attended and unattended sets of stimuli were interspersed with each other, distinguished only by their orthogonal orientations. In this arrangement, distractor items are also likely to have been captured when attention was spread over the target set.

These findings suggest that summary representations are generated at some point subsequent to the deployment of attentional resources. Nonetheless, in light of evidence that ensemble coding can circumvent capacity limitations in early visual processing, it remains possible that ensemble coding occurs prior to the registration of individual items in working memory. It is therefore important to isolate the stage at which this process first emerges.

The goal of the present study was to determine the locus of ensemble coding in relation to the allocation of attention and encoding in visual short-term memory. To do this, we used the "attentional blink" (AB) paradigm. The AB is a behavioural phenomenon that occurs under conditions of rapid serial visual presentation (RSVP), in which a stream of stimuli is presented at a rate of around 10 items per second (Raymond, Shapiro, & Arnell, 1992; Weischelgartner & Sperling, 1987). Embedded within this stream are two target stimuli distinguished from the distractors by featural (e.g., colour) or category-based information (e.g., letters amongst number distractors). The targets are separated by a varying "lag" (temporal offset within the stream). The central, robust finding is that participants often fail to report the second target (T2) if it follows the first target (T1) at a short lag (an offset of 2 to 6 items, or approximately 200-600 ms). The effect is not due to a perceptual deficit related to the T2 stimulus itself, because performance is unimpaired if participants are told to ignore T1 and report only T2. A variety of potential mechanisms have been invoked to explain this failure to report T2, such as a central processing bottleneck (Chun & Potter, 1995), delayed attentional reengagement (Wyble, Potter, Bowman, & Nieuwenstein, 2011), the temporary loss of control over an input filter tuned to target properties (Di Lollo, Kawahara, Ghorashi, & Enns, 2005), or the active shutting of an attentional gate following T1 processing (Olivers & Meeter, 2008; Raymond, Shapiro & Arnell, 1992). Evidence from neurophysiological studies suggests that the AB results from a suppression of the allocation of attention to T2. Sergent, Baillet, and Dehaene (2005) demonstrated that the N2 event-related potential (ERP) component, which is linked to attentional engagement with target stimuli (Folstein & van Petten, 2007), is reduced in T2 stimuli that are undetected, or "blinked." A paradigm in which ensemble coding is required for a display presented within the temporal window of the AB thus could be a powerful way to explore whether the generation of

summary representations is contingent on attentional engagement with the stimuli in the ensemble.

In a recent study, Joo, Shin, Chong, and Blake (2009) examined whether extraction of summary size information from sets of circles can occur during the AB. In their experiment, participants performed an RSVP task in which they identified a digit (T1) embedded within a stream of letter distractors. The T2 stimulus that followed consisted of two sets of circles, one on the left and one on the right side of the display, and participants were asked to estimate which set had the larger mean size. The authors found no significant difference between T2 accuracy in the dual-task condition and accuracy in a T2-only condition, in which T1 was ignored. On this basis, they concluded that the mean size estimate can be computed even when attention is limited. However, several aspects of the study complicate this interpretation. First, the circle displays presented as T2 were unmasked; the AB is typically reduced or even eliminated for unmasked stimuli (Giesbrecht & Di Lollo, 1998). Acknowledging this issue, Joo et al. performed a control experiment in which they used a random-dot noise image as a mask and again found little evidence of an AB. However, random-dot noise images are unlikely to make effective masks for simple geometrical shapes (indeed, this has been our own experience; also see Enns and Di Lollo, 2000). Second, Joo et al. employed very different types of stimuli and tasks for the two targets: discriminating digits from letters in the case of T1 and ensemble coding of sets of circles for T2. Yet previous research has found that the AB may not occur when T1 and T2 differentially tap into featural versus configural processing channels (Awh, Serences, Laurey, Daliwhal, van der Jagt, & Dassonville, 2004). Accordingly, the question of whether summary statistics can be successfully extracted when attentional deployment is disrupted is still without a clear answer. We addressed this question in an AB task employing similar stimuli (faces) for both T1 and T2.

The AB paradigm can also be used to examine whether ensemble coding of a set of items requires the individual items to be registered in working memory, or if it instead occurs prior to this step. The AB appears to be underpinned by the process of encoding and consolidating T1 into working memory (Bowman & Wyble, 2007). The timing of the AB, approximately 200 to 600 ms following the onset of T1, coincides with the latency of the P3 ERP component evoked by T1 (McArthur, Budd, & Michie, 1999), which is linked to the contextual updating of information in working memory (see Polich, 2007, for a review). More specifically, the AB is linked to the temporal overlap between the T1-evoked P3 and the attention-related N2 component produced by T2 (Sergent et al., 2005), suggesting that updating of working memory interferes with the attentional engagement of T2. Varying the task difficulty associated with T1 is also known to affect the magnitude of the subsequent AB (Dux & Harris, 2007; Elliott & Giesbrecht, 2015; Giesbrecht, Sy, & Elliott, 2007; Jolicœur, 1999; Wierda, Taatgen, van Rijn, & Martens, 2013); and this is reflected at the neural level, where concomitant increases in the amplitude of the P3 evoked by T1 are associated with an increase in the size of the subsequent AB (Martens, Elmallah, London, & Johnson, 2006). We compared the size of the AB induced by a T1 task requiring the ensemble coding of multiple items with that induced by estimating the same feature from a single item. This allowed us to gauge whether ensemble coding requires the individual items in an ensemble to be encoded into working memory or whether only the summary representation itself is encoded.

Current study

The purpose of this study was to utilise the AB paradigm to examine two questions. First, is the successful extraction of a summary representation affected by the availability of attentional resources (Experiment 1)? Second, does the formation of a summary representation necessitate encoding of the individual items in working memory (Experiment 2)? Two tasks were used in each experiment; both employed face stimuli. One task was a gender discrimination of a single face. The other task required participants to estimate the mean emotional expression of a set of four faces or the expression of a single face (in separate blocks of trials). The order of these tasks was reversed between the first two experiments, such that T1 required gender discrimination and T2 required emotion estimation in Experiment 1, and vice versa in Experiment 2. The T1 and T2 tasks were separated by either a short lag (within the typical AB window) or a long lag (beyond the typical AB window). Performance on the emotion-estimation task was measured as the difference between the reported expression and the actual expression. A mixture-modelling procedure was conducted to estimate the extent to which these response errors were attributable to guessing; and the difference in the proportion of non-guess responses between the short and long lags was used as our measure of the AB. This procedure also allowed a more accurate measurement of the precision with which participants performed the emotion-estimation task by removing the effect of such guesses from their distribution of responses. If ensemble coding requires attentional engagement with the stimuli constituting the ensemble, we should predict the presence of an AB in estimating the emotional expression of the set of four faces when these appear as the second target (T2) in Experiment 1. Alternatively, if extraction of summary statistics does not require access to attentional resources then performance should be equivalent at both short and long lags. Furthermore, if ensemble coding requires each face within the ensemble to be encoded into working memory before computing the summary representation, we would predict a larger T2 deficit induced by the estimation of the mean expression of four faces compared to a single face when these appear as the first target (T1) in Experiment 2. Finally, we

tested whether the results from these experiments might be attributable either to participants adopting a particular strategy that circumvents the need to engage in ensemble coding at all, or to differential demands on the deployment of spatial attention (Experiment 3).

Experiment 1

Methods

Participants

A total of 25 undergraduate students were recruited for Experiment 1 in exchange for course credit. Experimental procedures were approved by the Human Research Ethics Committee of the University of Sydney, and all participants gave their informed consent in writing.

Apparatus

The experiment was programmed and run using PsychoPy software (Peirce, 2007), operating on a PC under Windows 7 OS, and displayed on a 17″ Sony Trinitron CRT monitor (1280 × 960 pixels; 85 Hz refresh rate) at a viewing distance of ~57 cm.

Stimuli

Separate sets of grayscale face stimuli were used for the *Gender-discrimination* (T1) and *Emotion-estimation* (T2) tasks (Fig. 1). All faces, however, were identical in size, subtending 4.3° of visual angle in height and 3.5° in width.

The Gender-discrimination stimuli consisted of 12 female faces and 12 male faces (without facial hair) taken from an online database (Utrecht ECVP stimulus set; http://pics.psych.stir.ac.uk/2D_face_sets.htm). The faces were cropped tightly using an oval frame, thereby removing any hair cues, and converted to grayscale.

The face stimuli used in the Emotion-estimation task were identical to those used in Experiment 3 of Haberman, Harp, and Whitney (2009). Three emotional expressions (happy, sad and angry) of a single individual, taken from the Ekman gallery (Ekman & Friesen, 1976), formed the basis of the Emotion-estimation stimulus set used in all trials. Fifty morphs were created, using Morph 2.5 software, through linear interpolation between each pair of the original expressions. This resulted in 150 expressions on a circular scale from happy to sad to angry and back to happy (Haberman et al., 2009, for details). While we refer to the distance between adjacent morphed expressions along the distribution as a step of one "emotional unit," it should be noted that this does not necessarily correspond to a psychological unit of discriminability.

Different backward-masking stimuli were created for the Gender-discrimination and Emotion-estimation stimuli. Using MATLAB (Mathworks Inc.), each of the face stimuli was divided into 4 × 5 blocks. Each individual block was then rotated (either 0°, 90°, 180°, or 270°) and then randomly combined to produce new, scrambled faces. An oval frame was then used to crop the scrambled faces to an overall similar shape to the original face stimuli.

Gender-discrimination task (T1)

In the Gender discrimination task, a single face was presented in the centre of the screen. Faces were selected pseudo-randomly such that an equal number of female and male faces were presented in each of the Single and Ensemble conditions of the Emotion-estimation task.

Emotion-estimation task (T2)

On each trial, an emotional expression was selected at random from the entire stimulus set. In the Single condition, this

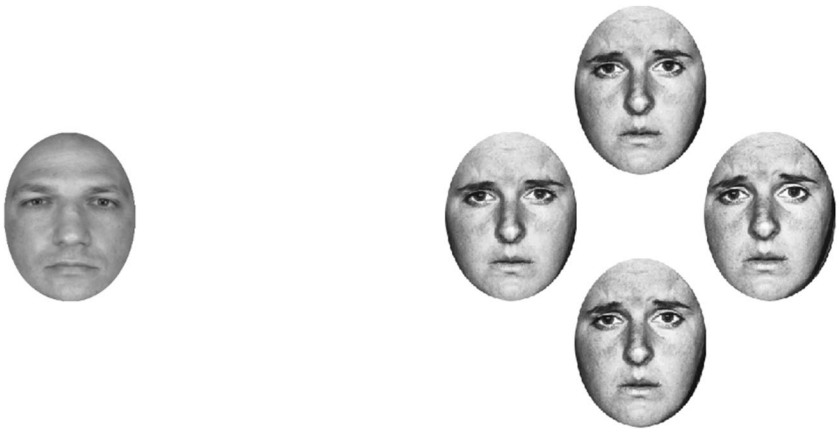

Fig. 1 Examples of stimuli used in the Gender-discrimination task (T1; left) and Emotion-estimation task (T2; right). Emotion-estimation stimuli were presented either in a set of four, arranged in a diamond configuration, or as a single face presented in the centre of the screen

expression alone was presented in the centre of the screen. In the Ensemble condition, this particular expression was not shown but instead provided the "mean" for a set of four emotional expressions that formed the ensemble. The four expressions selected were those 3 and 9 emotional units above and below the mean along the morph distribution (i.e., −9, −3, +3, and +9). Each expression thus differed from the others by at least 6 emotional units. Each face was then randomly allocated to a position at one of the points of an imaginary diamond. The centre of each face was 3.4° from the centre of the screen (Fig. 1), sufficient distance to avoid possible crowding effects (Whitney & Levi, 2011).

Procedure

Attentional-blink paradigms typically involve two target stimuli embedded in a stream of irrelevant distractor stimuli, which share some features with the targets. However, ensemble coding can integrate feature information over time as well as space (Haberman et al., 2009; Albrect & Scholl, 2010). Therefore, to minimize any potential influence from distractors, we used a "skeletal" RSVP design with no distractor stimuli—only targets and masks. This design still induces an attentional blink when the targets are presented with a relatively short stimulus onset asynchrony (SOA; Ward, Duncan, & Shapiro, 1997).

Each trial began with a black fixation cross, presented for 494–1000 ms against a white background (e.g., Fig. 2). This was replaced by a male or female face (T1) followed by a mask. In *Short-SOA* trials, the mask was immediately followed by the Emotion-estimation stimulus (T2) and subsequent mask. In *Long-SOA* trials, there was an additional 600-ms delay during which a fixation cross was visible, before the presentation of T2 and its respective mask. Based on pilot testing, stimuli for the Emotion-estimation task (in both the Single and Ensemble conditions) were displayed for 153 ms (13 monitor frames). Stimuli for the Gender-discrimination task, as well as mask stimuli for both tasks, were presented for 94 ms (8 monitor frames). This resulted in T1–T2 SOAs of 188 ms in the Short-SOA condition and 788 ms in the Long-SOA condition. In the Ensemble condition, the four faces displayed were followed by four masks presented in the same locations, whereas in the Single condition a single face was displayed centrally and followed by its mask. At the end of the trial, a blank screen was shown for 500 ms before participants were prompted to make their responses for the Gender-discrimination T1 task and then for the Emotion-estimation T2 task. No time limit was imposed on either response.

For the Gender-discrimination task, the screen prompt consisted of an upwards arrow pointing to the word "Woman" and a downwards arrow pointing to the word "Man." Participants pressed the up arrow or down arrow key to indicate whether they had seen a woman or a man, respectively.

For responses to the Emotion-estimation task, a probe expression, selected at random from the full set of 150 emotional expressions, was presented in the centre of the screen, with the text "Adjust the expression" below. Moving the mouse along its *x*-axis cycled the probe through the full distribution of expressions. Participants pressed the left mouse button to select the expression they thought best matched either that of the single expression (Single condition) or their estimate of the average of the four emotional expressions (Ensemble condition).

The Single and Ensemble conditions of the Emotion-estimation task were performed in separate blocks of 150 trials (75 each for the Short and Long SOA, randomly intermixed), with the order of the blocks counterbalanced between participants.

Analysis

Gender-discrimination task

Accuracy data were analysed using an ANOVA with SOA (Short vs. Long) and Condition (Single vs. Ensemble) as within-subjects factors.

Emotion-estimation task

We used a continuous response measure to assess participants' recall of the mean emotional expression. This is a procedure that has been employed previously to assess the precision of mean estimation (Albrecht & Scholl, 2010; Haberman et al., 2009). On each trial, response error was defined as the difference between the reported average and the true average on each trial. A Von Mises curve, which approximates a circular Gaussian distribution, was then fitted to the distribution of errors produced by each participant in each condition. We used a Von Mises curve because the distribution of errors is circular; for example, an error of +75 is indistinguishable from an error of −75. A participant's precision in estimating the mean is characterised by the dispersion of this fitted distribution. That is, we take the extent of trial-to-trial variability around the true mean to reflect the precision of the participant's internal representation of the mean. However, this is complicated by the presence of two different types of responses: guesses and non-guesses (or "informed" responses). While informed responses are expected to comprise a circular Gaussian distribution, guesses are drawn from a separate, uniform distribution. This is because the participant has no information about the stimulus that was presented, and must therefore select a response at random. Following Zhang and Luck's (2008) application of mixture modelling to a change-detection

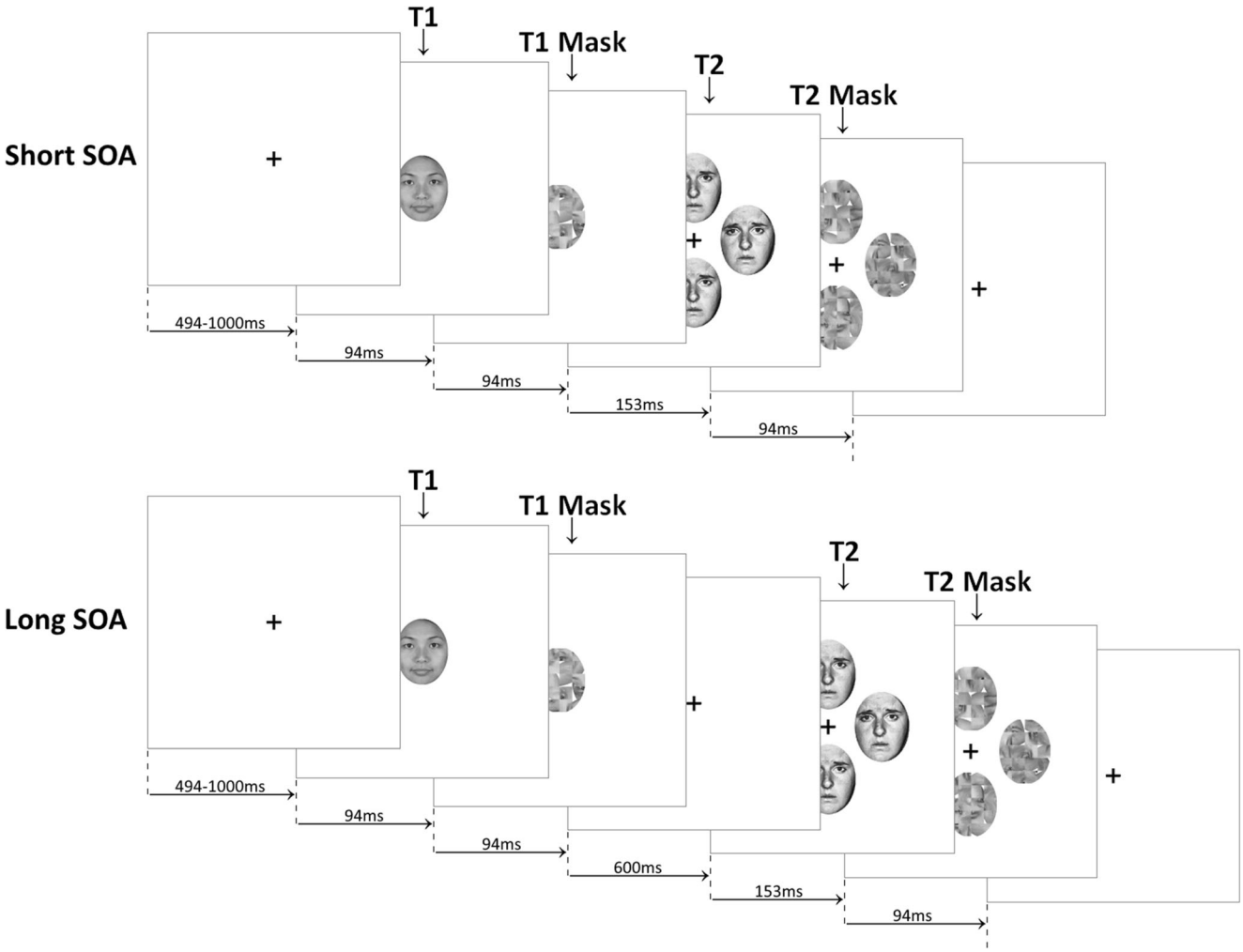

Fig. 2 Example of the trial procedure in Experiment 1. Both the Short-SOA (top) and Long-SOA (bottom) trials began with an initial fixation cross displayed for 494–1000 ms. Then, a male or female face (T1) was presented in the centre of the screen for 94 ms, before a backward mask was presented for a further 94 ms. In the Long SOA trials, the T1 mask was followed by a fixation cross for 600 ms. Then, in both types of trials, T2 was presented. T2 consisted of either four faces arranged in a diamond (Ensemble condition) or a single face in the centre of the screen (Single condition). The T2 stimulus was shown for 153 ms before a backward mask was presented for 94 ms. A fixation screen was then shown for a further 494 ms. Following the trial, participants were first prompted for their response to the Gender-discrimination T1 task and then for their response to the Emotion-estimation T2 task

paradigm, we adopted a procedure in which errors are assumed to reflect a mixture of a Gaussian distribution (informed responses) and a uniform distribution (guesses; see Asplund, Fougnie, Zughni, Martin, and Marois, 2014, for an example of this procedure applied to data from an AB paradigm). The relative contribution of each distribution to the fitted model can then be used to determine the probability that a participant is using information from the stimulus to make their response (P_e; see Fig. 3 for examples of these distributions from a representative participant). A lower probability of an informed response at a short lag compared to a long lag would then be indicative of an AB.

Only trials in which a correct response was recorded to the T1 stimulus (Gender-discrimination task) were included in the T2 analysis. The response-error values were converted to radians for analysis, however the reported results are in emotional units. Maximum-likelihood estimation was used to compute two parameters for analysis. The first was the relative weight of the von Mises curve in the final distribution (P_e), expressed as a percentage. This reflected the overall *efficacy* with which the participant produced informed responses (i.e., non-guesses) for that condition. A decrease in efficacy at the Short SOA compared to the Long SOA was taken as an indicator of an AB. We also calculated the concentration parameter (k) of the von Mises distribution, which was then converted to the standard deviation (σ) and used as our measure of the *precision* of the estimates from informed responses. (Note that this differs from the *temporal* precision parameter sometimes obtained from mixture modelling of response errors in the AB; see Goodbourn et al., 2016). A third parameter, the location (μ) of the von Mises distribution, was allowed to vary in the model but was not analysed. The

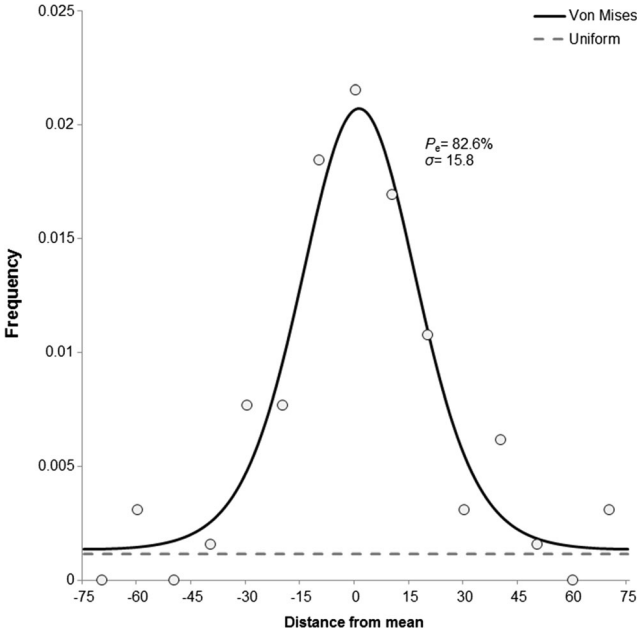

Fig. 3 Example of the fitting of a mixture of a uniform distribution (modelling "guesses") and a von Mises distribution (modelling non-guesses or "informed" responses) to the distribution of error responses (circles) from a representative participant. P_e = weighting of the von Mises distribution relative to the total distribution; σ = standard deviation of the von Mises distribution (in emotional units). To better illustrate the fit of the model, response errors are shown here in bins 10 emotional units wide; however, the fitting procedure was applied to the continuous response-error data

model was re-fit 400 times to each distribution, with starting parameters varying randomly ($0 \leq P_e \leq 1$; $-1 \leq \mu \leq 1$; $1 \leq k \leq 5$), to obtain the most likely model given the data. In order to prevent over-fitting, the most likely full model (uniform + von Mises) was formally compared to a restricted, uniform-only model using a likelihood-ratio test ($\alpha = .05$). When the uniform-only model was preferred, indicating that the participant never gave an informed response, efficacy was set to zero. Efficacy and precision were analysed separately in two-way, repeated-measures ANOVAs, with SOA (Short, Long) and Condition (Single, Ensemble) as within-subjects factors.

Results

Two participants had an overall T1 accuracy of less than 70 % on the Gender-discrimination task and were excluded from all analyses. A further four participants produced efficacy parameters of 0 at the Long SOA in the Ensemble condition, indicating that they were unable to perform this task under minimal task demands and were also excluded. This resulted in data from 19 participants being included in the analyses.

Gender-discrimination task (T1)

No differences in accuracy were found between Conditions (Single: 93.2 %, SEM = 0.7 %; Ensemble: 93.1 %, SEM = 0.8 %; $F_{(1,18)} < 1$) or between SOAs (Short: 92.6 %, SEM = 0.9 %; Long: 93.6 %, SEM = 0.8 %; $F_{(1,18)} = 1.66$, $p = .21$, $\eta_p^2 = .08$), nor was there a significant interaction between the two factors ($F_{(1,18)} = 1.29$, $p = .27$, $\eta_p^2 = .07$; Fig. 4).

Efficacy in Emotion-estimation task (T2)

The analysis revealed a significant main effect of Condition ($F_{(1,18)} = 13.34$, $p = .002$, $\eta_p^2 = .43$), with overall efficacy on Single-stimulus trials (81.2 %, SEM = 3.1 %) much higher than that on Ensemble-stimuli trials (61.4 %, SEM = 5.2 %). There was also a significant effect of SOA ($F_{(1,18)} = 18.69$, $p < .001$, $\eta_p^2 = .51$), with overall efficacy higher at the Long SOA (84.4 %, SEM = 2.5 %) compared with the Short SOA (58.2 %, SEM = 5.9 %). Finally, there was an interaction between Condition and SOA ($F_{(1,18)} = 6.6$, $p = .019$, $\eta_p^2 = .27$). The effect of SOA was greater for Ensemble trials (Short SOA: 41.1 %, SEM = 9.1 %; Long SOA: 81.8 %, SEM = 4.0 %; $p < .001$) than that for Single trials (Short SOA: 75.3 %, SEM = 6.1 %; Long SOA: 87.1 %, SEM = 2.4 %; $p = .11$). The higher efficacy on Single trials relative to Ensemble trials was only observed at the Short SOA ($p = .003$), with no significant difference at the Long SOA ($p = .23$; Fig. 5).

Precision in Emotion-estimation task (T2)

The analysis of the precision data indicated a main effect of Condition ($F_{(1,18)} = 16.3$, $p = .001$, $\eta_p^2 = .48$), with greater precision for Single stimuli ($\sigma = 17.3$, SEM = 1.2) than for Ensemble stimuli ($\sigma = 25.5$, SEM = 1.8). Overall, precision was higher at the Long SOA ($\sigma = 17.4$, SEM = 1.0) compared

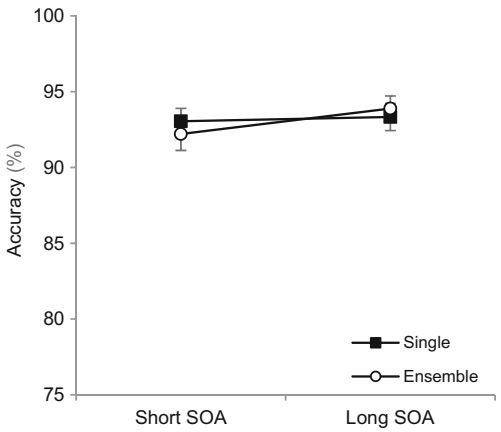

Fig. 4 Mean T1 Accuracy at Short and Long SOAs for Single and Ensemble conditions in Experiment 1. Error bars indicate ±1 SEM

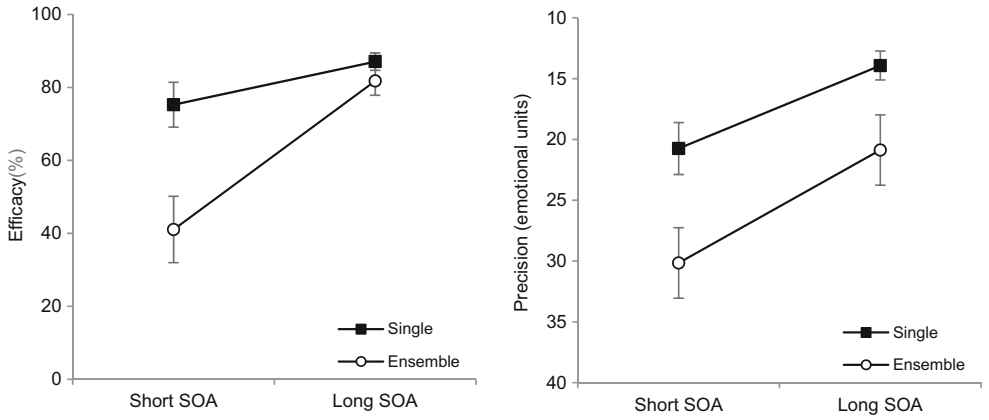

Fig. 5 Mean T2 Efficacy (left) and T2 Precision (right) scores at Short and Long SOAs for Single and Ensemble conditions in Experiment 1. Note that the *y*-axis is reversed in the right panel such that higher positions correspond to better precision (lower standard deviation). Error bars indicate ±1 SEM

with the Short SOA ($\sigma = 25.5$, SEM = 1.99; $F_{(1,18)} = 13.9$, $p = .002$, $\eta_p^2 = .44$). The interaction between Condition and SOA was not significant ($F_{(1,18)} < 1$).

Discussion

Consistent with the idea that extracting summary statistics occurs *after* attentional resources have engaged with the items within the ensemble, these results indicate that ensemble coding is affected by disruptions to the allocation of attention to stimuli. This is in contrast to the findings of Joo et al. (2009), who reported that computing the mean size of a set of stimuli was not affected by the AB. This discrepancy might be due to differences in featural processing between the stimuli used for T1 and T2, and a lack of effective backward masking of T2 in the paradigm employed by Joo et al. Both of these factors may have reduced the likelihood of eliciting an effective AB (Awh et al., 2004; Giesbrecht & Di Lollo, 1998).

In addition, we observed a significantly *greater* AB deficit in the Ensemble condition than in the Single-stimulus condition. This suggests that the ensemble-coding process itself imposes an additional cost on central processing resources, and therefore that summary representations may not be computed automatically. This conflicts with previous studies reporting very little or no effect of concurrent task demands on the accuracy of such representations (Chong & Treisman, 2005a). One note of caution is that we did not find a significant difference between efficacy at the Short and Long SOAs for the Single condition. This may be because we used face stimuli, which have been shown in a number of studies to be less affected by the AB than other types of stimuli (Awh et al., 2004; Jackson & Raymond, 2006; Landau & Bentin, 2008). While Awh et al. (2004) found that face stimuli could experience a substantial AB when they were used as the two targets (as in the present study), the discrepancy may be due to the difficulty of the face-identification task they employed at T1. Landau and Bentin (2008) found that the magnitude of the AB suffered by faces at T2 was modulated by the difficulty of a racial-discrimination task performed on faces presented at T1. When the T1 task was easy (yielding approximately 82 % accuracy), the magnitude of the induced AB was relatively small (approximately 5 %). In our experiment, the T1 gender-discrimination task yielded very high accuracy (93 %), which may explain why we failed to observe a significant AB effect in the Single condition.

Alternatively, however, the relatively better performance in the Single condition at the Short SOA may arise because T2 appears in the same central spatial location as T1, thus not necessitating any shifts of attention. In comparison, the Ensemble condition requires a broadening of the attentional spotlight to take in the four expressions presented at more peripheral locations. Spatial attention is believed to freeze, or even shrink, during the AB (Dell'Acqua, Sessa, Jolicœur, & Robitaille, 2006; Olivers, 2004), and this may result in fundamental differences in the ability to deploy attention in the two conditions. This issue is addressed more directly in Experiment 3.

As well as a decline in efficacy during ensemble coding, we also found that the precision of mean estimates was worse at Short SOAs than at Long SOAs. This was true for both Single and Ensemble stimuli. Moreover, the precision of mean estimates was worse for Ensemble than for Single stimuli. While this demonstrates that attention is required for accurate representation of local (single-item) information, these results also support the conclusion that accurate ensemble coding relies critically on the availability of attentional resources. In contrast, many other studies on summary statistics generally report summary estimates to be about as good as estimates for single stimuli (Ariely, 2001; Chong & Treisman, 2003). This raises the question of whether our finding could be attributed to observers adopting a nonoptimal strategy when performing the task in the Ensemble condition. For example, because of the reduction in attentional resources available for T2, they

may be encoding only one expression, rather than attempting to average all of the expressions. Such a subsampling strategy could be reasonably successful in the present task, because selecting any single expression provides a fairly close approximation of the mean. However, the resulting distribution of errors would have a higher variance (i.e., lower precision) than in the Single condition, because the selected expression would range between −9 and +9 emotional units from the mean across trials. We investigate this possibility in Experiment 3.

Having established a lower bound on the processing stage associated with the emergence of summary representations (i.e., after attentional engagement) we next turn to examining the relationship between ensemble coding and working memory. To do this, in Experiment 2 we reversed the order of the T1 and T2 stimuli and presented the Emotion-estimation task at T1. If estimating the mean of an Ensemble display requires each item to be encoded into working memory, we would expect to see a greater AB effect on the subsequent Gender-discrimination T2 task following the Ensemble condition compared with the Single-stimulus condition.

Experiment 2

Methods

Participants

A total of 24 undergraduate students took part in Experiment 2, none of whom had participated in Experiment 1. All participants gave their informed consent in writing.

Stimuli and procedure

The stimuli, tasks, and procedures used in Experiment 2 were identical to those used in Experiment 1, apart from reversing the order of tasks: The Emotion-estimation task was presented as T1, and the Gender-discrimination task as T2. Because different durations were used for Emotion-estimation and Gender-discrimination stimuli, this resulted in slightly different latencies for the T1–T2 SOAs in Experiment 2: The Short SOA was 247 ms, whereas the Long SOA was 847 ms.

Results and discussion

The data for Experiment 2 were analysed in the same manner as those for Experiment 1. Three participants demonstrated extremely poor performance for the Emotion-estimation task at T1, as indicated by efficacy parameters of 0, and were excluded. This resulted in data from 21 participants being analysed for Experiment 2.

Efficacy in Emotion-estimation task (T1)

We found no effect on efficacy of Condition (Single: 78.1 %, SEM = 3.8 %; Ensemble: 78.6 %, SEM = 3.9 %; $F_{(1,20)} < 1$) or SOA (Short: 78.7 %, SEM = 3.6 %; Long: 78.0 %, SEM = 4.1 %; $F_{(1,20)} < 1$); nor was the interaction between the two factors significant ($F_{(1,20)} = 2.82$, $p = .11$, $\eta_p^2 = .12$; Fig. 6).

Precision in Emotion-estimation task (T1)

Estimates were more precise in the Single condition ($\sigma = 16.1$, SEM = 1.1) compared with the Ensemble condition ($\sigma = 22.6$, SEM= 1.3; $F_{(1,20)} = 17.92$, $p < .001$, $\eta^2 = .47$). Precision did not vary with SOA (Short: $\sigma = 20.1$, SEM = 1.0; Long: $\sigma = 18.6$, SEM = 1.3; $F_{(1,20)} < 1$), nor was there an interaction between Condition and SOA ($F_{(1,20)} < 1$; Fig. 6).

Gender-discrimination task (T2)

All T2 trials were included in the analysis. Given the continuous nature of the T1 measurement, T2 data were not conditionalised on T1 performance. We observed no effect on T2 Gender discrimination of the type of stimulus encoded in the Emotion-estimation task at T1 (Single: 79.3 %, SEM = 1.9 %; Ensemble: 77.5 %, SEM = 1.9 %; $F_{(1,20)} = 1.63$, $p = .22$, $\eta_p^2 = .08$). There was an effect of SOA ($F_{(1,20)} = 118.38$, $p < .001$, $\eta_p^2 = .85$) with lower accuracy at the Short SOA (70.0 %, SEM = 2.3 %) relative to the Long SOA (86.8 %, SEM = 1.6 %). There was no interaction between Condition and SOA ($F_{(1,20)} < 1$; Fig. 7).

Discussion

As may be expected given that the T1 task could receive full attention, efficacy (the proportion of informed responses) was approximately the same in both Ensemble and Single stimuli trials. However, the precision of informed responses was worse on Ensemble trials. Indeed, precision was remarkably consistent with that observed at the Long SOA in Experiment 1 for both Single and Ensemble trials (approximately 15 and 20 emotional units, respectively). Despite the increased amount of attention that could be directed towards the Emotion-estimation task, efficacy did not reach 100 % for either condition. This likely results from the sheer difficulty of the task, involving the delayed recall of brief, backward-masked, complex visual stimuli and is also reflected in the very sizeable AB induced for the second target. Importantly, however, we observed no difference in the magnitude of the AB produced in the Gender-discrimination task by Ensemble and Single stimuli, indicating similar working-memory demands in both conditions. This is consistent with the notion

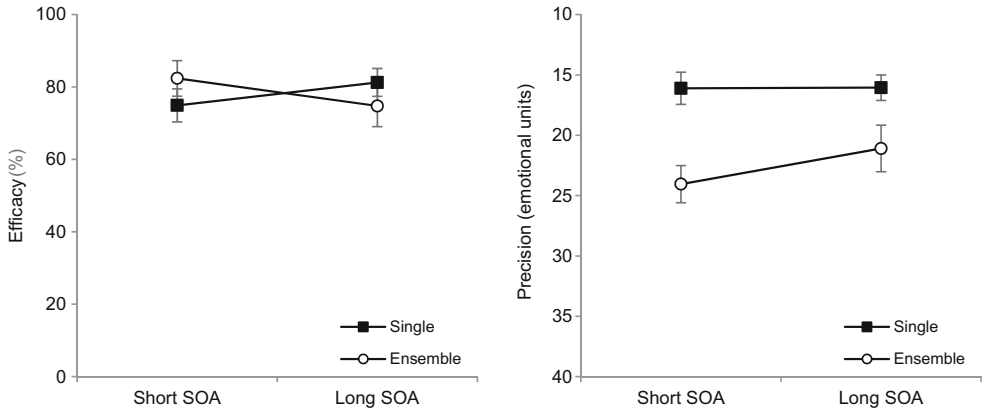

Fig. 6 Mean T1 Efficacy (left) and T1 Precision (right) at Short and Long SOAs for Single and Ensemble conditions in Experiment 2. Note that the *y*-axis is reversed in the right panel such that higher positions correspond to better precision (lower standard deviation). Error bars indicate ±1 SEM

that computing summary representations does not require individual items to be encoded in working memory.

As in Experiment 1, we found that the precision of the emotion estimation was considerably lower for the Ensemble than for the Single condition, even though attentional resources are presumably fully engaged with the task. This raises some alternative explanations that may account for this result. Firstly, it is possible that observers encode all items in working memory but these representations are very low-resolution. This could result in imprecise estimates of the mean expression while at the same time consuming similar working-memory resources as a good-quality single representation and, consequently, inducing a similar-sized AB for the T2 task. Although we cannot rule this out using the current set of data, we consider this explanation unlikely. Haberman and Whitney (2007) have shown that observers do not seem to have access to information about individual faces in similar displays of four stimuli, even when the stimuli are presented for extended periods (2 s) and thus have ample opportunity to be encoded in visual working memory. Another possible explanation is that the lower precision in the Ensemble condition may once again be the result of a subsampling strategy. If participants were instead basing their judgements on a single expression and encoding this in working memory, this might also result in an AB deficit of a magnitude equivalent to that in the Single condition. As we have suggested previously, such a strategy might be encouraged by the high similarity between the four expressions presented in the ensemble, because choosing any one of these expressions could have provided a reasonable approximation of the mean. It is important to exclude this alternative before interpreting the present results as reflecting the cognitive demands of ensemble averaging. This was the aim of Experiment 3.

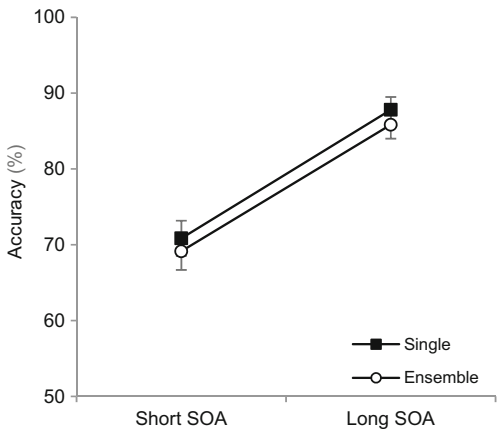

Fig. 7 Mean T2 Accuracy on the Gender-discrimination task at Short and Long SOAs for Single and Ensemble T1 conditions in Experiment 2. Error bars indicate ±1 SEM

Experiment 3

Experiment 3 was based on Experiment 1 with the following changes. First, we introduced an additional condition in which participants were explicitly instructed to subsample from a set of stimuli. This condition was similar to the Ensemble condition in that a set of four facial expressions was presented, but here participants were required to focus on and report the emotional expression of only one face. Comparing precision in this condition with precision in the Ensemble condition should reveal whether the results of the Ensemble condition could be explained by subsampling. This is because calculating the error of responses in the Subsampling condition relative to the mean of the entire ensemble, rather than the specific target expression, provides a simulation of participants employing such a strategy in the Ensemble condition. If this were the case, then precision in the Subsampling condition (using error measured relative to the mean of the ensemble) should be the same as in the Ensemble condition. If, on the other hand, participants use more than one expression in their estimates, the process of averaging would result in higher precision in the Ensemble condition.

Second, the variability of the emotional expressions in the set was increased, in order to reduce the similarity between individual expressions and the mean of the ensemble. We

expected this change would discourage reliance on a subsampling strategy. However, it was important that stimuli were not too dissimilar, because this might prevent ensemble coding from occurring at all (Utochkin & Tiurina, 2014). Accordingly, we doubled the distance between each member of the set from 6 emotional units to 12.

Finally, we introduced the requirement for a spatial shift of attention in the Single condition, by presenting the face away from fixation. All three conditions of this experiment thus required participants to shift attention beyond the location where the T1 face was presented.

Methods

Participants

Thirty-six undergraduate students took part in Experiment 3. None had participated in either Experiment 1 or Experiment 2. All participants gave their informed consent in writing.

Stimuli and procedure

The stimuli and procedure used in Experiment 3 were identical to those used in Experiment 1, unless otherwise mentioned. There were three conditions (rather than two) in the Emotion-estimation task at T2. The Ensemble condition was identical in all respects, other than the increased emotional distance between stimuli, to that in Experiment 1. The Single condition differed from that in Experiment 1 in that the face was presented 3.4° above the centre of the screen (i.e., the position of the topmost face in the Ensemble condition; Fig. 8, left panel). In the additional Subsampling condition, the display was identical to the Ensemble condition, with four faces presented in a diamond arrangement. However, participants were instructed to attend to only the top face amongst the four. For the Subsampling and Ensemble conditions, the distance in emotional units between individual expressions in the set was increased from 6 to 12. This meant that the four expressions were spaced −18, −6, +6, and +18 emotional units from the trial mean (Fig. 8, right panel, for an example of such a set). Finally, the gender-discrimination response for T1 was recoded to use the left and right arrow keys rather than the up and down keys, in case any upward shift in attention toward the target face caused a response bias. The three tasks were completed in separate blocks of 75 trials each (Short- and Long-SOA trials randomly intermixed), with the order of the blocks randomised across participants.

Analysis

The data for Experiment 3 were analysed in the same manner as in Experiments 1 and 2 but with the addition of Subsampling to the Condition factor (Single, Subsampling, Ensemble). As a result, a Huynh-Feldt correction was applied to correct for possible violations of sphericity, although we report the uncorrected degrees of freedom for convenience.

In the initial analysis, responses in the Subsampling condition were evaluated against the expression of the target face (i.e., how accurately and precisely the participant estimated the attended emotional expression). We conducted a further analysis in which responses on the Subsampling task were treated as estimations of the mean of the entire ensemble, rather than estimates of the particular face to which they were

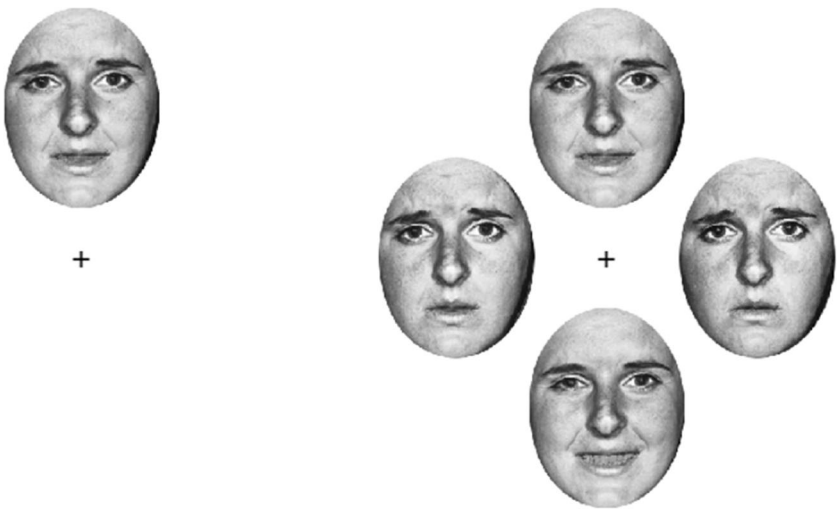

Fig. 8 Examples of stimulus configurations used in the Emotion-estimation task (T2) in Experiment 3. In the Single condition (left panel), one emotional expression was presented above fixation. In the Subsampling and Ensemble conditions (right panel), stimuli were presented in a set of four, arranged in a diamond configuration

instructed to attend. This analysis provided a direct analogue for measurements that would be obtained in the Ensemble condition if participants adopted the strategy of attending to only one expression in the set. We called this modelled data set $Subsample_{Mean}$. The efficacy and precision parameters estimated from these error distributions were directly compared to those generated in the Ensemble condition using a within-subjects ANOVA with Condition ($Subsample_{Mean}$ vs. Ensemble) and SOA (short vs. long) as its factors to test whether performance in the Ensemble condition is distinguishable from that generated by an explicit subsampling strategy.

Results

Six participants were excluded because of extremely poor performance at the Long SOA, as indicated by efficacy parameters of 0. Data from 30 participants were analysed for Experiment 3.

Gender-discrimination task (T1)

There were no differences in accuracy between Conditions (Single: 90.3 %, SEM = 1.3 %; Subsample: 91.6 %, SEM = 1.1 %; Ensemble: 90.2 %, SEM = 1.1 %; $F_{(2,58)} = 2.07$, $p = .14$, $\eta_p^2 = .07$). However, accuracy was significantly worse on Short-SOA trials (89.4 %, SEM = 1.1 %) compared with Long-SOA trials (92.1 %, SEM = 1.1 %; $F_{(1,29)} = 21.25$, $p < .001$, $\eta_p^2 = .42$; Fig. 9). There was also a significant interaction between the two factors ($F_{(2,58)} = 3.68$, $p = .03$, $\eta_p^2 = .11$). At the Short SOA, there were no differences between the three conditions (Single: 89.2 %, SEM = 1.3 %; Subsample: 89.4 %, SEM = 1.2 %; Ensemble: 89.6 %, SEM = 1.2 %; $p = .97$). But at the Long SOA, Gender-discrimination performance was better on Subsample trials (93.9 %, SEM = 1.1 %)

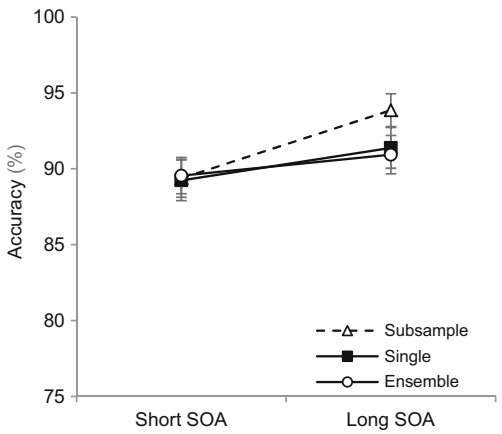

Fig. 9 Mean T1 Accuracy at Short and Long SOAs for Single, Subsample, and Ensemble conditions in Experiment 3. Error bars indicate ±1 SEM

than on both Single (91.4 %, SEM = 1.3 %; $p = .015$) and Ensemble (90.9 %, SEM = 1.3 %; $p = .009$) trials. The Single and Ensemble conditions did not differ from each other ($p = .98$).

Efficacy in Emotion-estimation task (T2)

As in Experiment 1, only trials in which T1 was correct were included in the analysis. There was no difference in efficacy between the three conditions (Single: 60.3 %, SEM = 3.7 %; Subsample: 63.8 %, SEM = 4.0 %; Ensemble: 64.0 %, SEM = 4.5 %; $F_{(2,58)} < 1$). Efficacy on Short-SOA trials (44.6 %, SEM = 5.41) was significantly lower than on Long-SOA trials (80.7 %, SEM = 2.2; $F_{(1,29)} = 43.33$, $p < .001$, $\eta_p^2 = .60$), indicating a pervasive AB. This effect of SOA did not interact with Condition ($F_{(2,58)} < 1$; Fig. 10).

Precision in Emotion-estimation task (T2)

There was an overall effect of Condition on the precision of the Emotion-estimation task ($F_{(2,58)} = 18.15$, $p < .001$, $\eta_p^2 = .38$). Estimates of emotional expression were less precise in the Ensemble condition ($\sigma = 29.2$, SEM = 1.1) compared with the Single ($\sigma = 22.5$, SEM = 1.6; $p = .005$) and Subsampling conditions ($\sigma = 19.4$, SEM = 1.5; $p < .001$), in which only one face was attended, whereas the latter two did not differ significantly from each other ($p = .138$). Precision was worse at the Short SOA ($\sigma = 29.2$, SEM = 1.6) compared with the Long SOA ($\sigma = 18.1$, SEM = 0.8) across all three conditions ($F_{(1,29)} = 51.51$, $p < .001$, $\eta_p^2 = .640$). There was no interaction between the two factors ($F_{(2,58)} < 1$; Fig. 10).

Efficacy in emotion-estimation task (T2): $Subsample_{Mean}$ versus Ensemble

Efficacy was higher in the $Subsample_{Mean}$ condition (74.6 %, SEM = 4.5 %) than in the Ensemble condition (64 %, SEM = 4.5 %; $F_{(1,29)} = 5.36$, $p = .03$, $\eta_p^2 = .16$; Fig. 11). This is contrary to the result expected if participants were performing the Ensemble condition by subsampling only one face. Overall efficacy was lower at the Short SOA (52.1 %, SEM = 6.9 %) compared with the Long SOA (86.5 %, SEM = 2.2 %; $F_{(1,29)} = 26.22$, $p < .001$, $\eta_p^2 = .48$), indicating an AB. There was no significant interaction between SOA and Condition ($F_{(1,29)} = 2.79$, $p = .11$, $\eta_p^2 = .09$).

Precision in Emotion-estimation task (T2): $Subsample_{Mean}$ versus Ensemble

The precision of responses was significantly higher for $Subsample_{Mean}$ ($\sigma = 26.7$, SEM = 1.0) compared with the Ensemble condition ($\sigma = 29.2$, SEM = 1.1; $F_{(1,29)} = 4.69$, $p = .04$, $\eta_p^2 = .14$). Precision was also overall lower at the Short

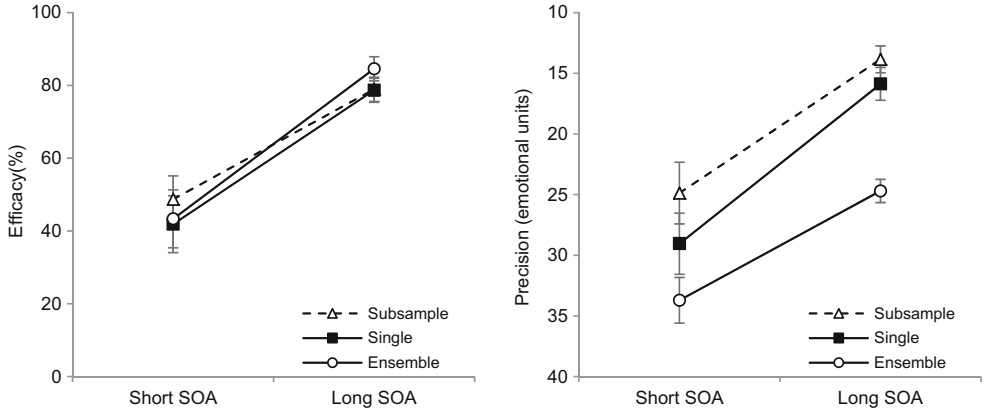

Fig. 10 Mean T2 Efficacy (left) and T2 Precision (right) at Short and Long SOAs for Single, Subsample, and Ensemble conditions in Experiment 3. Note that the *y*-axis is reversed in the right panel such that higher positions correspond to better precision (lower standard deviation). Error bars indicate ±1 SEM

SOA (σ = 32.6, SEM = 1.6) relative to the Long SOA (σ = 23.3, SEM = 0.7; $F_{(1,29)} = 32.44$, $p < .001$, $\eta_p^2 = .53$), and this did not vary across conditions ($F_{(1,29)} < 1$; Fig. 11).

Discussion

The goal of Experiment 3 was to address some of the outstanding questions that arose from Experiments 1 and 2. The first was whether the systematically poorer performance in the Ensemble condition, compared with the Single condition, could be the result of a subsampling strategy. The results obtained in the Subsampling condition, in which participants were explicitly directed to focus on only one expression, clearly indicate that this is not the case. The data from this condition were subjected to an analysis in which response error was measured relative to the mean of the entire ensemble, rather than the single face on which participants were instructed to base their responses. This simulated a subsampling strategy being employed on ensemble stimuli. Using this approach, we found significant differences between the Subsample and Ensemble conditions in both the efficacy and precision of mean estimation, inconsistent with the use of a common strategy across the two conditions. We expected to observe equivalent performance in the Ensemble condition if observers' judgements were based on only a single item in the display. Instead, participants were *less* successful and had *worse* precision in the Ensemble condition than in the Subsampling condition. Thus, the present results indicate that attempting to spread attention over an entire array of stimuli has a deleterious effect on summarising the featural properties of those objects. Together with our finding of worse efficacy and precision at short SOAs than at long SOAs, this reinforces the conclusion that ensemble coding is affected by impaired allocation of attention.

The second question that we addressed was whether a key finding in Experiment 1—specifically that ensemble coding was more severely affected by the AB than was single-expression coding—could be attributed to differences in attentional-shifting requirements. In the present experiment, where both the Single and the Ensemble conditions required a shift in spatial attention from the central location of the T1

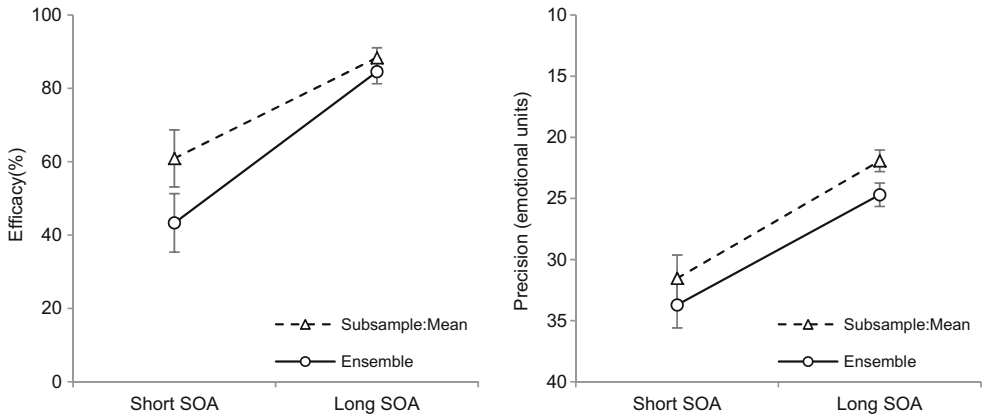

Fig. 11 Mean T2 Efficacy (left) and T2 Precision (right) at Short and Long SOAs for $\text{Subsample}_{\text{Mean}}$ and Ensemble conditions in Experiment 3. Note that the *y*-axis is reversed in the right panel such that higher positions correspond to better precision (lower standard deviation). Error bars indicate ±1 SEM

stimulus, the difference between Ensemble and Single conditions was no longer present. Performance in both conditions was affected by the AB to a similar extent, in terms of both the efficacy and the precision of informed responses. This finding indicates that ensemble coding does not impose any additional cost on attentional resources over and above estimating the emotional expression of a single face. However, the presence of an AB does suggest that ensemble coding depends on the deployment of spatial attention across the set of stimuli.

An additional finding was that T1 accuracy was lower when the subsequent T2 stimulus occurred at a short SOA compared with a long SOA. This effect is sometimes observed in AB paradigms and is thought to reflect competition between T1 and T2 for attentional resources and encoding into working memory (Potter, Staub, & O'Connor, 2002; Hommel & Akyürek, 2005). While such an effect was not found in Experiment 1, this competition may have been exacerbated by the additional requirement to shift attention from the location of T1 in the Single and Subsample conditions. Performance on the T1 task also differed between the three experimental conditions at the long SOA, but not when they were separated by a short SOA. At present, we have no satisfactory explanation for this latter finding, but we note that it does not affect any of our central conclusions.

General discussion

The current study employed an attentional-blink paradigm to investigate two questions about the processing of visual summary statistics. The first was whether ensemble coding of featural information (facial expression) is affected by the availability of attentional resources. The second was whether forming a summary representation requires each item in the set to be first encoded into working memory.

The first question was addressed in Experiment 1. Participants reported the average expression of a set of faces presented as the second task in a dual-task RSVP paradigm, either during or after the expected temporal window for the AB. The key finding of this experiment was that the Ensemble condition suffered an AB, as indexed by a reduction in efficacy (reflecting increased guessing) at Short SOA compared with Long SOA. In contrast to the earlier study by Joo et al. (2009), these results demonstrated that limiting the attentional resources available to process an ensemble impairs the representation of summary statistics. In addition, efficacy and precision in this Ensemble condition were compared to a control condition in which the expression from a single face was reported. The AB observed for the Ensemble condition was significantly larger than that for the Single condition. This may suggest that ensemble coding is even more sensitive to the depletion of attentional resources than is encoding a single face; alternatively, it may be attributable to differences between the two conditions in the spatial re-allocation of attention after the T1 task. When this latter aspect of the procedure was equated in Experiment 3, the magnitude of the AB was similar between the two conditions, supporting the second alternative. We also observed a reduction in the precision of estimates for informed responses (non-guesses) at the Short SOA compared with the Long SOA in both Experiments 1 and 3. Because the effect of guesses is removed by the modelling process, this may reflect the fact that even when they make informed responses participants sometimes have only limited control of attention following encoding of T1. Attention may not be deployed effectively across the entire ensemble at the Short SOA, thereby reducing the accuracy of the mean estimation.

The second question of whether extracting a summary representation from a set of stimuli requires all items to first be encoded into working memory was addressed in Experiment 2. Participants performed the Emotion-estimation task as the first in the sequence. We compared the size of the AB in a subsequent T2 task (determining the gender of a single face) under these two conditions. The key finding of this experiment is that single stimuli and ensemble sets of stimuli produced a near-identical AB in the subsequent T2 task. This indicated that summary representations are formed without the need to encode the individual items into working memory. If this were necessary, we should have observed a larger AB in the Ensemble condition. The findings of these experiments indicated that ensemble coding requires attentional engagement but occurs prior to the consolidation of information in working memory.

These results can be interpreted within Treisman's (2006) framework for the role of attention in visual perception. According to this model, which builds on Feature Integration Theory (Treisman & Gelade, 1980), feature maps represent the distribution of a particular feature (e.g., the colour red) across a visual scene. Spatial information within the feature maps is retained only implicitly via the topographical organisation of the low-level units that feed this information forward. These feature maps in turn indiscriminately activate the object representations with which they are associated. When attention is focused on a particular object within the scene, this boosts activity from those feature maps fed by units processing information from its spatial location, which acts to bind these features together into a stable object representation. When attention is distributed across multiple objects, the representation of all of them is boosted, and feature maps become dominated by those objects. However, without focused attention to promote individuation of each of the items, only the summaries from the feature maps are accessible for encoding in working memory (and, consequently, subsequent retrieval). In other words, summary representations are formed automatically from objects within the spatial distribution of attention through compulsory pooling of their featural information.

In our study, the process of consolidating T1 into working memory impairs the deployment of attentional resources to the following T2 stimulus. In Experiment 1, this causes a greater reduction in efficacy for ensemble stimuli relative to a single stimulus. This is because the Ensemble condition requires a broadening of spatial attention from the location of T1 to spread over all items in the subsequent T2 stimulus set. In comparison, no such shift is required in the Single condition. Owing to limitations in the allocation of attention imposed by the AB, this prevents the construction of a feature map for emotional expression more often in the Ensemble condition than in conditions where only a single stimulus must be attended. The observation of a reduction in precision at the short SOA may also reflect a situation where the impaired control of attention affects the quality of the feature map produced for the T2 stimulus. The results from Experiment 2 further suggest that the representations derived from these feature maps, whether originating from a single stimulus or summarised from ensemble sets of stimuli, have a fixed cost on working memory. This is consistent with a recent study demonstrating that contralateral delay activity, an EEG measure of working-memory load, is consistent across multiple set sizes during ensemble coding (Baijal, Nakatani, van Leeuwen, & Srinivasan, 2013). Finally, in Experiment 3 we equated the need to shift attention from the location of T1 between the Single and Ensemble conditions. We observed no difference in the magnitude of the AB between the two conditions. This supports the idea that the extraction of summary statistics does not impose any additional processing costs compared with a single stimulus (i.e., occurs automatically) provided that spatial attention can be effectively deployed (Chong & Treisman, 2005a).

In Experiments 1 and 2, the precision of the mean estimates for ensemble sets was significantly poorer than the precision of estimates for a single face. This raised the possibility that participants may have used a subsampling strategy to perform the Emotion-estimation task in the Ensemble condition. If subjects had performed the task simply by selecting a single face from the ensemble display, we would expect to observe systematically lower precision compared with the single-face condition. Some authors have suggested that ensemble coding may actually reflect exactly this kind of strategic behaviour, such as basing size judgments on only the largest or smallest items in the display (Myczek & Simons, 2008; Marchant, Simons, & de Fockert, 2013). Such a strategy could also have been encouraged by the similarity of the faces we used in the ensemble sets in those experiments. However, the results of Experiment 3 suggest otherwise. In that experiment, we reduced the similarity between the faces in the displays by increasing their separation in emotional-expression space. We also directly compared performance in the Ensemble condition with performance in a condition in which the participants were explicitly instructed to focus on a single face from the set, and found that the precision in the Ensemble condition was in fact poorer than the precision of estimating the expression of a single face. This demonstrates that our results cannot be explained by participants adopting such a subsampling strategy. Rather, they suggest that disrupting attentional deployment impairs the ability to summarize the statistics of an ensemble of stimuli.

The superior precision for the Subsample condition compared to the Ensemble condition does raise the question of exactly why a single face chosen from the ensemble would provide a more precise estimate of the mean than averaging across all of the faces in the ensemble. This is particularly puzzling given that ensemble coding has been shown previously to provide accurate estimates of the average emotional expression of sets of faces with displays similar to those used here (Haberman & Whitney, 2009). It is possible that the task demands of the AB paradigm prevented participants from engaging in ensemble coding at all. However, if so, it is unclear exactly what they could be doing instead. Participants are not simply guessing, as our mixture-modelling procedure removes the effect of guessing from estimates of precision. The results of Experiment 3 also exclude subsampling of a single face as an alternate strategy. Furthermore, the results of Experiment 2 are inconsistent with serially encoding a subset of items (more than one, but less than the full ensemble) and mentally averaging these; and in any case, such a strategy would be expected to provide a *better* estimate of the mean than subsampling a single face, rather than a worse estimate as we observed in Experiment 3. Thus, we believe that the most likely explanation is that our participants are attempting to perform ensemble coding, but their attempts result in poor estimates of the mean. This is likely due to the brief presentation and the backward masking of stimuli in our experiments, as well as the depletion of attentional resources inherent in the AB. While average size can be estimated with reasonable accuracy in as little as 50–100 ms, the accuracy of such estimates has been shown to improve with prolonged stimulus duration (Chong & Treisman, 2003); and complex stimuli, such as faces, may require even more time for accurate extraction of summary statistics. Previous studies have shown that mean estimates of emotional expression can be computed at stimulus durations below 200 ms, but their accuracy is comparatively poor (Haberman & Whitney, 2009). Under appropriate circumstances, a summary representation computed via ensemble coding might provide less reliable estimates of the mean than a single face within the set. Nonetheless, these results indicate that the extraction of average emotional expression is detrimentally affected when access to attentional resources is restricted during the AB.

Our study focused on one particular type of ensemble coding: estimating the average emotion of a set of faces. This process has been shown to exhibit many of the same properties as averaging across other features, such as size and

orientation (Haberman & Whitney, 2009). Summary information derived from faces likely taps into configural processing, given that inverted and scrambled faces do not show the same level of accuracy in ensemble coding as upright faces (Haberman et al., 2009). Because configural processing is likely to introduce additional attentional demands, it is possible that other properties may be more resilient to the disruption of attention than are faces and their emotional expressions. This may explain some of the differences between our findings and those of Joo et al. (2009) and could be investigated in future studies.

It is worth noting that the current set of experiments required the formation and retention of a rather precise representation of the mean to perform the adjustment task at the end of the trial. Therefore, it is still possible that summary representations may be computed preattentively, as suggested by various studies (Oriet & Brand, 2013), but that attention is required for maintenance and explicit access to this information in a similar way to what is observed with some Gestalt grouping processes (Kimchi & Razpurker-Apfeld, 2004). This could be investigated in future studies by looking at how summary statistics derived implicitly (e.g., from distractors that are outside the focus of attention) interact with the type of attentional bottlenecks investigated.

In conclusion, the findings from the present study indicate that summary representations of the emotional content of sets of faces necessitate the deployment of attention across the ensemble without requiring the individual items to be registered in working memory. This accords with a model in which information about a particular feature set is pooled automatically across the full spatial extent of attention prior to the encoding of information into working memory. Given that the distribution of attention can be intentionally controlled, some of the inconsistencies in the ensemble-coding literature may also arise because subjects adopt their own idiosyncratic strategies for performing an experimental task.

Acknowledgments This research was supported by a Discovery Project Grant (DP120102299) and a Future Fellowship (FT0992123) from the Australian Research Council awarded to I. M. Harris.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*(4), 560–567.

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83,* 25–39.

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research, 101,* 34–40.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W., & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological Science, 25*(3), 824–831.

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1440–1449.

Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive Psychology, 48,* 95–126.

Baijal, S., Nakatani, C., van Leeuwen, C., & Srinivasan, N. (2013). Processing statistics: An examination of focused and distributed attention using event related potentials. *Vision Research, 85,* 20–25.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review, 114*(1), 38–70.

Chakravarthi, R., & Cavanagh, P. (2009). Recovery of a crowded object by masking the flankers: Determining the locus of feature integration. *Journal of Vision, 9*(10), 4.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*(1), 1–13.

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45,* 891–900.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 109–127.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138,* 289–301.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology, 62*(9), 1716–1722.

Dell'Acqua, R., Sessa, P., Jolicœur, P., & Robitaille, N. (2006). Spatial attention freezes during the attention blink. *Psychophysiology, 43,* 394–400.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*(4), 481–507.

Di Lollo, V., Kawahara, J., Ghorashi, S. M. S., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research, 69,* 191–200.

Dux, P. E., & Harris, I. M. (2007). Viewpoint costs arise during consolidation: Evidence from the attentional blink. *Cognition, 104,* 47–58.

Dux, P. E., Visser, T. A. W., Goodhew, S. C., & Lipp, O. V. (2010). Delayed re-entrant processing impairs visual awareness: An object substitution masking study. *Psychological Science, 21,* 1242–1247.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Elliott, J. C., & Giesbrecht, B. (2015). Distractor suppression when attention fails: Behavioral evidence for a flexible selective attention mechanism. *PLoS ONE, 10*(4), e0126203.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking. *Trends in Cognitive Sciences, 4*(9), 345–352.

Folstein, J. R., & van Petten, C. (2007). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45,* 152–170.

Giesbrecht, B., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1454–1466.

Giesbrecht, B., Sy, J. L., & Elliott, J. C. (2007). Electrophysiology evidence for both perceptual and postperceptual selection during the attentional blink. *Journal of Cognitive Neuroscience, 19*(12), 2005–2018.

Goodbourn, P. T., Martini, P., Barnett-Cowan, M., Harris, I. M., Livesey, E. J., & Holcombe, A. O. (2016). Reconsidering temporal selection in the attentional blink. *Psychological Science, 27*(8), 1146–1156.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from set of faces. *Current Biology, 17*(17), R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718–734.

Hommel, B., & Akyürek, E. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *Quarterly Journal of Experimental Psychology, 58*(8), 1415–1433.

Huang, L. (2015). Statistical properties demand as much attention as object features. *PLoS ONE, 10*(8), e0131191.

Im, & Halberda. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception & Psychophysics, 75,* 278–286.

Jackson, M. C., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics, 68,* 543–557.

Jacoby, O., Kamke, M. R., & Mattingly, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 233–244.

Jolicœur, P. (1999). Concurrent response-selection demands modulate the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1097–1113.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision, 9*(9), 7.

Kimchi, R., & Razpurker-Apfeld, I. (2004). Perceptual grouping and attention: Not all groupings are equal. *Psychonomic Bulletin & Review, 11*(4), 687–696.

Landau, A. N., & Bentin, S. (2008). Attentional and perceptual factors affecting the attentional blink for faces and objects. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 818–830.

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142,* 245–250.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences, 9*(6), 296–305.

Martens, S., Elmallah, K., London, R., & Johnson, A. (2006). Cuing and stimulus probability effects on the P3 and the AB. *Acta Psychologica, 123*(3), 204–218.

McArthur, G., Budd, T., & Michie, P. (1999). The attentional blink and P300. *NeuroReport, 10,* 3691–3695.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70,* 772–788.

Olivers, C. N. L. (2004). Blink and shrink: The effect of the attentional blink on spatial processing. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 613–631.

Olivers, C. N. L., & Meeter, M. (2008). Boost and bounce theory of temporal attention. *Psychological Review, 115*(4), 836–863.

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research, 79,* 8–16.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially liable. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1149–1162.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849–860.

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11*(12), 1–8.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391–1400.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*(4-8), 411–443.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and de Fockert. *Acta Psychologica, 146,* 7–18.

Ward, R., Duncan, J., & Shapiro, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual presentation. *Perception & Psychophysics, 59*(4), 593–600.

Weischelgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238*(4828), 778–780.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science, 15*(4), 160–168.

Wierda, S. M., Taatgen, van Rijn, & Martens. (2013). Word frequency and the attentional blink: The effects of target difficulty on retrieval and consolidation processes. *PLoS ONE, 8*(9), e73415.

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General, 140*(3), 488–505.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.