Ensemble coding of face identity is not independent of the coding of individual identity

Markus F. Neumann, Ryan Ng, Gillian Rhodes, and Romina Palermo

The University of Western Australia

Author Note

Markus F. Neumann, Ryan Ng, Romina Palermo, and Gillian Rhodes, ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, The University of Western Australia, WA, Australia

This research was supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021) and an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300).

Correspondence concerning this article should be addressed to Markus Neumann, School of Psychology, The University of Western Australia; 35 Stirling Highway, Crawley, WA 6009, Australia. E-mail: markus.neumann@uwa.edu.au

Abstract

Information about a group of similar objects can be summarized into a compressed code, known as ensemble coding. Ensemble coding of simple stimuli (e.g., groups of circles) can occur in the absence of detailed exemplar coding, suggesting dissociable processes. Here, we investigate whether a dissociation would still be apparent when coding facial identity, where individual exemplar information is much more important. We examined whether ensemble coding can occur when exemplar coding is difficult, as a result of large sets or short viewing times, or whether the two types of coding are positively associated. We found a positive association, whereby both ensemble and exemplar coding were reduced for larger groups and shorter viewing times. There was no evidence for ensemble coding in the absence of exemplar coding. At longer presentation times, there was an unexpected dissociation, where exemplar coding increased yet ensemble coding decreased, suggesting that robust information about face identity might suppress ensemble coding. Thus, for face identity, we did not find the classic dissociation - of access to ensemble information in the absence of detailed exemplar information - that has been used to support claims of distinct mechanisms for ensemble and exemplar coding.

Keywords: set averaging, ensemble coding, ensemble representation, crowd, face identity

Word count: 5676 (excluding title page, abstract, and references)

Introduction

Human visual processing resources are limited. As an example, the capacity of visual working memory is approximately four items (e.g., Luck & Vogel, 1997). Nevertheless, our visual experience of the environment is rich and detailed (Oliva & Torralba, 2006), suggesting that our visual impressions are illusory or that the limitations of the visual system may be able to be overcome. One way the visual system may overcome limitations might be via summarizing information about similar items into a compressed code, such as the average property of the group. These averages or "ensemble representations" have been shown for the visual processing of various types of information, such as the size of geometric shapes (e.g., Ariely, 2001; Chong & Treisman, 2003, 2005b), the orientation of textures (Dakin & Watt, 1997), and the direction of motion of dots (Watamaniuk, Sekuler, & Williams, 1989). Ensemble coding has also been shown for more complex stimuli, such as facial expressions (e.g., Haberman & Whitney, 2007, 2009, 2010), facial gender (Haberman & Whitney, 2007), and the direction of eye gaze (Sweeny & Whitney, 2014), as well as in the perception of tonal sequences in the auditory domain (Piazza, Sweeny, Wessel, Silver, & Whitney, 2013).

It has been argued that ensemble coding might be a qualitatively distinct process to that used for coding of individual exemplars (e.g., Chong & Treisman, 2005b; Corbett & Melcher, 2014; Whitney, Haberman, & Sweeny, 2014). Three pieces of evidence are consistent with the idea of distinct processes underlying ensemble and exemplar coding. First, studies have manipulated the number of items in a group to examine whether the extraction of ensemble information from a group is a serial process, by which participants successively focus attention on individual items, or a "global" process, operating in parallel across the whole group at once (e.g., Robitaille & Harris, 2011). These studies have often found that an increase in set size is not detrimental to the precision of the ensemble representation (Ariely, 2001; Attarha, Moore, & Vecera, 2014; Chong & Treisman, 2005b; Haberman & Whitney, 2007), while information about individual exemplars decreases with increased set size (Robitaille & Harris, 2011). For instance, Haberman and Whitney (2009) reported that accuracy in determining the location of an individual emotional face in sets of 1, 2, 3, or 4 faces declined from perfect accuracy for set size of 1 to about 50% at a set size of 4 (Experiment 4a), and exemplar identification performance in a two-alternative forced-choice task dropped from about 90% for set size of 1 to about 65% for a set size of 3 (Experiment 4b), while increasing set size from 4 to 16 faces had no effect on the precision of the ensemble representation (Experiment 1).

Second, ensemble coding has been characterized as a very rapid process. Ensemble representations can be formed from sets that are seen for only about 50 ms (Chong & Treisman, 2005b; Haberman & Whitney, 2009). In contrast, coding of exemplars within a group takes longer – at least 500 ms to determine the size of an individual circle (Ariely, 2001) and 2 s when judging the expression or identity for a face seen in a group (Haberman & Whitney, 2007). Third, ensemble information may be available even when precise individual exemplar information is not. For example, ensemble perception has been demonstrated when individual gabor patches were crowded such that participants were unable to report their orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), for items in the neglected hemifield of neglect patients (Pavlovskaya, Soroker, Bonneh, & Hochstein, 2015), and for items for which participants experienced change blindness (Haberman & Whitney, 2011).

A central argument across these different studies is that ensemble coding must be, to some extent, dissociable from the coding of individual exemplars. If ensemble coding is an early implicit feed-forward process, as suggested by Hochstein and colleagues (Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015), an experimental manipulation that impairs exemplar coding (such as reducing encoding time, or crowding individual items) may not reduce ensemble coding to the same extent. It is important to note though that demonstrations of ensemble coding despite poor representations for individual exemplars do not indicate that individual exemplars were never coded. An alternative possibility is that the exemplars were averaged into an ensemble representation and subsequently discarded (Alvarez, 2011). The discarding of individual exemplar representations might be especially likely for information that is of no immediate relevance to the perceiver. For example, it is not obvious why participants should remember sizes of individual circles from a group, particularly if the task requires participants to code the “mean” size of the group, rather than the sizes of individual circles. Even for more complex objects, individual exemplars might not be as important as the ensemble property: Individual trees in a forest, single blades of grass in a field, or single cars in a parking lot may not matter to us (unless of course we are searching for our car in the parking lot), and might therefore not be stored with individuating information (Chong & Treisman, 2005a).

However, there are stimulus classes, for which the coding of individuating information is important, and may also be more likely to be retained, most notably the identity of an individual face. Correctly identifying a person facilitates social functioning, and the face is the primary source of information about a person’s identity (O'Toole et al., 2011). Humans can encode very stable representations of a person’s face that allows them to identify the person accurately, even under impoverished viewing conditions (Jenkins & Burton, 2011). Therefore, for groups of faces, it seems likely that if individual identity information is associated with the coding of the ensemble identity, then this individual information could also be retained. There is some support for this proposal from studies showing ensemble coding with relatively preserved coding of individual identities (de Fockert & Gautrey, 2013; Kramer, Ritchie, & Burton, 2015; Neumann, Schweinberger, & Burton, 2013). Moreover, a link between ensemble and exemplar coding of face identity was found in a recent individual differences study (Haberman, Brady, & Alvarez, 2015), with a strong correlation between the coding of face identity exemplars and ensembles, $r(45) = .76$. However, all of these studies presented small groups of four faces at a relatively long duration of 1-2 s, allowing ample opportunity to individuate each of the four faces.

In the present study, we manipulated difficulty of encoding individual exemplars to examine whether ensemble representations of face identity can be dissociated from exemplar representations of the individual identities in a group. We utilized experimental manipulations used previously to address this question with less complex objects (as described earlier), specifically: altering the number of faces in the group (set size, Experiment 1) and the presentation duration of the group (Experiment 2). If ensemble coding for face identity can be dissociated from exemplar coding, then we would expect to see the type of dissociation found in previous studies (e.g., Ariely, 2001; Haberman & Whitney, 2009, 2011; Parkes et al., 2001), whereby ensemble representations are present even when individual exemplar recognition is reduced (for larger set sizes and shorter presentation durations). Alternatively, if ensemble coding for face identity is closely associated with the individuation of individual face exemplars as suggested by a recent study (Haberman et al., 2015), then we would expect a reduction of ensemble coding when exemplar recognition is impaired.

To encourage the encoding of exemplar representations for individual faces even when this was difficult (that is, for large groups and short presentations), we used a member identification paradigm, in which participants viewed groups of faces and then indicated whether or not a single subsequently presented probe face was part of the preceding group (de Fockert & Gautrey, 2013; de Fockert & Wolfenstein, 2009; Kramer et al., 2015; Neumann et al., 2013; Rhodes, Neumann, Ewing, & Palermo, 2015). The probe face could be a previously seen face (exemplar), or a computationally derived morphed image between the faces in the group, i.e., the *set average*. Given that the set averages were never presented in the group, they should not be endorsed as ‘seen’ and when this occurs it is evidence for ensemble coding of faces (de Fockert & Gautrey, 2013). This approach has been widely used to assess both explicit memory for individual exemplars of a group, and incidental memory for ensemble information. For example, Ariely (2001) has used the member identification paradigm to demonstrate the power of ensemble coding by showing that even when participants were asked to encode individual exemplars, they were very sensitive to the mean property of a group: they endorsed the circle of the mean size of a previously presented group as seen even though the mean had not been an actual member of the group. The member identification paradigm is well-suited for addressing the relationship between ensemble and exemplar coding for several reasons. First, any reduction in individual exemplar coding is very likely the result of the experimental manipulation, and is not a failure of coding individual exemplars as a result of task instructions that do not require participants to remember the individual exemplars. Second, the member identification task allows testing (implicit) ensemble coding and (explicit) exemplar coding within the same task. Lastly, implicit ensemble coding as tested in a member identification task is a good approximation of how ensemble information could influence people’s decisions in natural settings, such as identifying a face that has been seen in the context of a group.

Experiment 1: Individual exemplar and ensemble representations of face identity with variation in set size

In this experiment we manipulated set size (number of faces in the group). We predicted that increasing set size would be associated with a reduction in accurate recognition of individual face exemplars seen as part of a group. The novel question is whether increasing set size would preserve, or even facilitate (Robitaille & Harris, 2011), strength of ensemble identity coding, which would mean that the two processes are dissociable.

Methods

Participants.

Twenty-four Caucasian participants (17 females, mean age = 18.6, SD = 0.92) participated. All had normal or corrected-to-normal vision and gave informed consent. The study was conducted in compliance with guidelines approved by the University of Western Australia's Human Research Ethics Office.

Stimuli.

Sixty-four male Caucasian face images were obtained from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010) and the CAL/PAL face database (Minear & Park, 2004). All faces were of neutral expression. Photoshop CS6 (Adobe Systems, San Diego, CA, USA) was used to transform the images into greyscale, remove facial hair, and adjust the images to similar luminance levels. All faces were embedded in a mask that covered ears and hair. Eight faces of characters from the Cartoon series "The Simpsons" were additionally sourced from the Internet and used as example stimuli during practice trials.

These faces were used to create 64 "sets", with 16 sets for each of four set size levels - "2", "4", "6", and "8". Each face was approximately 2.6° x 3.2° (width, height) visual angle (VA) at a viewing distance of 90 cm when presented in a group, and 3.4° x 4.3° VA when presented as single probe faces (cf. Figure 1). The dimensions of the largest set ("8") were approximately 12.8° x 15.5° VA. Generally, each of the individual faces appeared in several different sets, but never more than once in the same set. The faces in set size "4" were positioned within a two by two matrix, and those in sets of size "2" occupied two adjacent positions within this matrix. For size "8" sets, additional faces were positioned on the left, right, top and bottom of the two by two matrix. Set size "6" utilised only the left/right or top/bottom positions (see example in Figure 1). Sixty-four "set averages" were created by averaging the shape and texture information of the individual faces in each set, using the face mixer tool in FantaMorph 5.4.1 (http://www.abrosoft.com).

Procedure.

The experiment was controlled by E-prime 2 software (Psychology Software Tools, Inc, Sharpsburg, PA, USA). Participants initiated a trial by pressing the space bar. The trial began with the presentation of a central digit, specifying the number of faces in the following set. A set was then presented for 1600 ms, followed immediately by a single probe face for 500 ms duration. Participants were then prompted to respond by a blank screen with a reminder of the question ("Present in previous group?"). Participants indicated whether or not they thought that the probe image had been presented in the previous set, by pressing one of the keys "a" and "l" on a standard keyboard, which were labelled "yes" and "no", respectively.

Probes appeared centrally on screen (to avoid retinal match of set and probe images) and corresponded to one of four different probe types: (i) Set average match: the set average created from averaging previous set faces, (ii) Set average mismatch: a set average created from averaging faces that were not in the previous set, but contained the same number of faces, (iii) Exemplar match: a face exemplar presented in the previous set; or (iv) Exemplar mismatch: a face exemplar not presented in the set. Example probe types are illustrated in Figure 1.

The experiment included 256 trials in total (16 per probe type for each of the 4 set size levels), presented in random order. Breaks were allowed after every 64 trials. The experiment took about 25 minutes to complete.

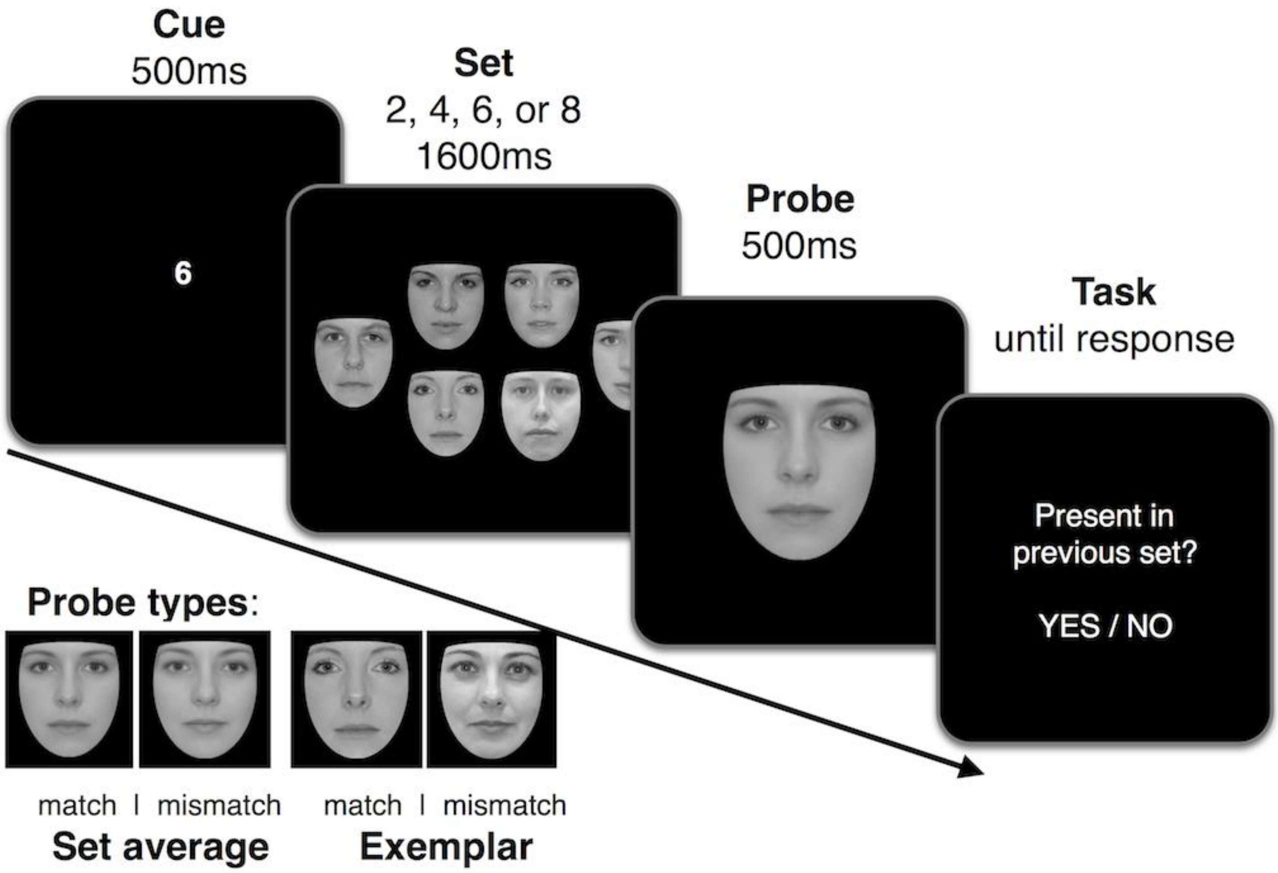

Figure 1. Illustration of trial procedure in Experiment 1. Each trial started with a cue indicating the number of faces in the following set display. The set contained faces from different identities, followed by a single probe. When prompted by the task screen, participants had to indicate whether or not they believed they had seen the probe face in the preceding set. Probe faces could be morphs between set faces (Set average match), or between an equivalent number of unseen faces (Set average mismatch), or could be one face of the set (Exemplar match) or an unseen face (Exemplar mismatch).

Data analysis.

Table 1 displays the proportions of “present” responses for all conditions. Following Rhodes and colleagues (2015), we calculated “endorsement scores” by subtracting mismatch from match conditions separately for exemplar and average probe types. Endorsement scores are well suited to assess the relationship between ensemble and exemplar coding of identity, because they reflect participants' sensitivity to probes that contain identity information from the previous set while eliminating potential biases that are not of interest for the present research. Endorsement scores are displayed in Figure 2. They were analysed in a within-subjects ANOVAs with the factors of Probe Type (exemplar, set average) and Set Size (2, 4, 6, 8).

Table 1. Mean proportions of "Present" responses for all probe types as a function of set size. Values in parentheses give standard errors of the means.

	Set size			
Probe Type	2	4	6	8
Exemplar match	.84 (.02)	.58 (.03)	.52 (.03)	.50 (.02)
Exemplar mismatch	.15 (.03)	.31 (.03)	.32 (.03)	.42 (.02)
Set Average match	.77 (.03)	.48 (.03)	.39 (.03)	.28 (.03)
Set Average mismatch	.18 (.02)	.27 (.03)	.24 (.03)	.28 (.04)

Results and Discussion

Results are depicted in Figure 2, and were clear-cut: with increasing set size, endorsement scores were reduced both for set averages and individual exemplars. For individual exemplars, this decrease was expected, and reflects reduced accuracy to identify a probe face as study set size increases. Participants were also less likely to falsely endorse a set average as seen when set size increased, indicating reduced strength of the ensemble representation with increased set size. These observations were confirmed by a repeated measures ANOVA, revealing a main effect of Probe Type, $F(1,23) = 9.73$, $p = .005$, $\eta_P^2 = .297$, with larger endorsement scores for exemplar ($M = .307$, $SE = .019$) than set average ($M = .238$, $SE = .016$) probes. The main effect of Set Size was also significant, $F(1,23) = 198.25$, $p < .001$, $\eta_P^2 = .896$, and Bonferroni-corrected contrasts between set size levels indicated that endorsement scores significantly declined with each increase in set size level (all *p*s < .01). Importantly, the interaction of Probe Type and Set Size was not significant,

$F(1,23) = 0.16$, $p = .925$, $\eta_P^2 = .078$, offering no evidence for a dissociation between the coding of exemplars and set averages for face identity.

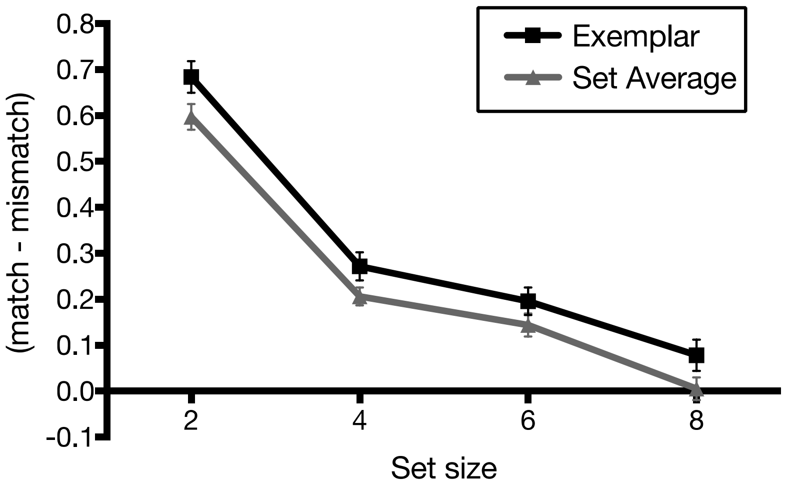

Figure 2. Endorsement scores (match minus mismatch) as a function of set size. Error bars indicate SEM.

Experiment 2: Individual and ensemble representations with variation in set duration

To further explore the relationship between coding of individual exemplar and ensemble information, we examined the time course of these processes in Experiment 2, by manipulating the presentation duration of the sets. All sets contained four faces, as ensemble representations reliably occur for sets of four faces (de Fockert & Gautrey, 2013; de Fockert & Wolfenstein, 2009; Neumann et al., 2013). Restricting presentation time should decrease accuracy with which individual exemplars can be encoded. However, if ensemble coding is dissociable from coding of individual exemplars, then ensemble representations could still be coded even when groups were only very briefly seen.

Method

Participants.

Twenty-four Caucasian participants (17 female, mean age = 21.3, *SD* = 4.9) were recruited who had not participated in Experiment 1. All had normal or corrected-to-normal vision and gave informed consent. The study was conducted in compliance with guidelines approved by the University of Western Australia's Human Research Ethics Office.

Stimuli.

The same faces from Experiment 1 were used, and sets were assembled in the same manner, except that all sets contained four faces.

Procedure.

Faces were randomly positioned in the 2 x 2 matrix and shown for each of eight set durations. Trials of the same duration were grouped into "short" (50, 100, or 200 ms), "medium" (400, 800, or 1600 ms), and "long" (3200 or 6400 ms) blocks. Block order was counterbalanced across participants, such that 1/3 of the participants started with a short, medium, or long block, respectively[1]. Within these blocks, trials of each duration were presented together (as in Haberman & Whitney, 2009, Experiment 2). The order of durations within each block was random. To familiarize participants with the upcoming set durations, each of the eight durations started with two practice trials. In addition to these 16 practice trials, there were a total of 512 experimental trials (16 per probe type for each of the eight set duration levels). The whole experiment took approximately 40 minutes to complete.

Results and Discussion

The proportions of "present" responses for all conditions are shown in Table 2. They were used to calculate unbiased recognition scores as in Experiment 1. As expected, endorsement scores

[1] We blocked the presentation durations into short, medium and long blocks so that we could examine whether different block orders would encourage systematic differences in participants' strategies, or influence processing on subsequent blocks in other ways (e.g., facilitating performance on "short" blocks when having seen "long" blocks first, for instance as a result of priming). To foreshadow, we found no such block order effects.

for exemplars increased with increasing set duration (Figure 3). A very similar pattern was found for endorsement scores of set averages up to 1600 ms. At long durations (3200 and 6400 ms), however, endorsement scores for set averages declined, whereas they continued to increase for exemplars.

Table 2. Mean proportions of “present“ responses for all probe types as a function of presentation duration. Values in parentheses give standard errors of the means.

	Presentation duration							
Probe Type	50	100	200	400	800	1600	3200	6400
Exemplar match	.39 (.04)	.36 (.03)	.42 (.03)	.48 (.04)	.58 (.03)	.66 (.04)	.70 (.04)	.79 (.03)
Exemplar mismatch	.35 (.04)	.34 (.04)	.40 (.05)	.32 (.04)	.33 (.04)	.28 (.05)	.24 (.03)	.23 (.04)
Set Average match	.48 (.06)	.51 (.05)	.49 (.05)	.38 (.04)	.50 (.05)	.57 (.05)	.45 (.05)	.45 (.06)
Set Average mismatch	.47 (.06)	.43 (.05)	.39 (.04)	.28 (.04)	.30 (.05)	.24 (.04)	.26 (.04)	.28 (.04)

An initial ANOVA with Probe Type and Presentation Duration as within-subject factors, and Block Order (short, medium, long block seen first) as between-subject factor revealed no main effect of Block Order, and no interaction with the other factors, all $F < 2.5$, all $p > .10$. Thus, Block Order was omitted from the following analyses.

An ANOVA with repeated measures factors of Probe Type (Exemplar, Average) and Presentation Duration (8 levels: 50, 100, 200, 400, 800, 1600, 3200 ms) revealed main effects of Probe Type, $F(1,23) = 31.31$, $p < .001$, $\eta_P^2 = .577$, Presentation Duration, $F(7,161) = 34.90$, $p < .001$, $\eta_P^2 = .603$, and an interaction, $F(7,161) = 11.34$, $p < .001$, $\eta_P^2 = .330$ (Figure 3). Polynomial contrasts showed that across the full continuum endorsement scores linearly decreased with reduced presentation duration, $F(1,23) = 214.6$, $p < .001$, $\eta_P^2 = .903$, indicating overall lower performance at shorter durations.

To examine the relationship between exemplar and ensemble representations, we calculated pair-wise t-tests to compare set average and exemplar recognition scores on each level of set duration. No difference was found between exemplar and set average endorsement at 50, 100 or 200 ms durations, all *t*s < 1.5, all *p*s > .15. Bonferroni-corrected paired-sample t-tests indicated that none of the exemplar or set average endorsements for either of these durations were significantly different from chance (0), all t's < 2.5, all p's > .05, except for set average endorsements at the 200 ms duration, $t(23) = 3.43$, $p = .032$ (note that set average endorsements at the 200 ms duration were not significantly differerent from exemplar endorsements). This indicates individuation of faces in a group was difficult at short presentation durations, and that set averaging was similarly difficult. No dissociation between the two processes is incompatible with the idea that ensemble representations can be extracted from very briefly presented groups, when exemplar coding is impaired (Haberman & Whitney, 2009). Moreover, set average and exemplar endorsement scores did not differ over the 400 and 1600 ms range either, all *t*s < 1.5, all *p*s > .15. During this time range, a substantial increase in endorsements of ensembles was observed when representations for the individual face exemplars became more accurate. Bonferroni-corrected paired-sample t-tests indicated that all exemplar and set average endorsements were still not different from chance at 400 ms duration, both t's < 3, both p's > .05, whereas all endorsements were significantly different from chance for 800 and 1600 ms durations, all t's > 4.9, all p's < .001.

However, endorsement scores for set averages were significantly lower than for exemplars at the longer durations of 3200 and 6400 ms: $t(23) = 5.49$, $p < .001$, Cohen's $d = 1.13$; and $t(23) = 9.99$, $p < .001$, Cohen's $d = 2.05$, respectively. That is, endorsement scores for set averages increased only until 1600 ms and then significantly declined between 1600 and 6400 ms, $t(23) = 3.54$, $p = .002$, Cohen's $d = 0.731$. In contrast, endorsement scores for exemplars increased beyond 1600 ms, $t(23) = 4.72$, $p < .001$, Cohen's $d = 0.964$. This pattern is suggestive of a dissociation between ensemble and exemplar coding of face identity, albeit not the expected one. It suggests that ensemble representations may be displaced by individual exemplar representations when these become sufficiently detailed.

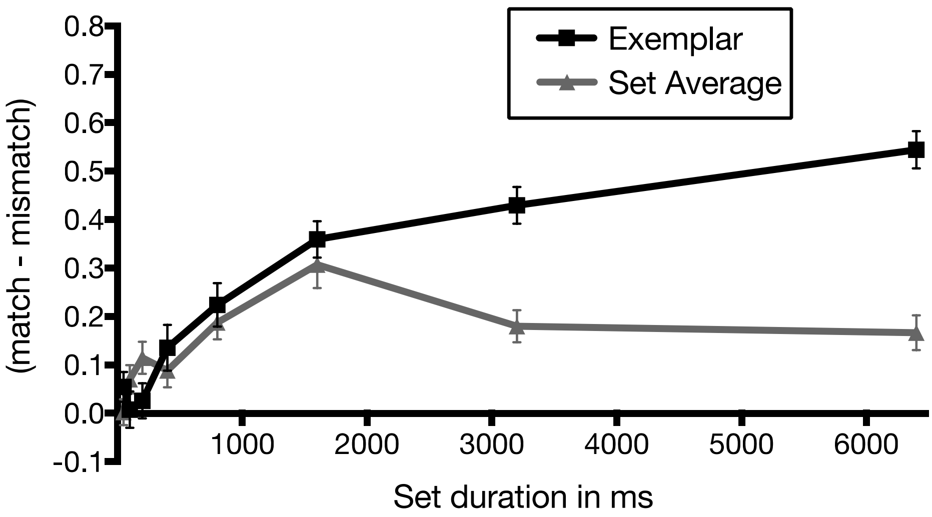

Figure 3. Endorsement scores (match minus mismatch) as a function of set duration in Experiment 2. SEMs are shown.

General Discussion

In this study, we manipulated the difficulty of encoding individual exemplars by varying set size and presentation duration, and these manipulations affected the recognition of exemplars as expected. Our novel question was to examine whether ensemble coding of identity would be similarly affected by these manipulations, suggesting a relationship between the two kinds of processes or whether ensemble coding would still occur, when individual recognition was disrupted. The data from both experiments suggest that there is a relationship between the two processes. When individual identities were difficult to recognise, because either the set size was large or faces were seen briefly, then there was also no evidence for preserved ensemble identity coding. Rather, in Experiment 1, smaller set sizes that showed more accurate exemplar identity coding also showed more ensemble coding. In Experiment 2, ensemble identity coding increased, as did exemplar identity coding, when faces were seen for longer durations (up to 1600 ms) yet there was no evidence for differential reduction of the two types of coding at very brief durations. Taken together, these results indicate a close link between the coding of ensemble and exemplar identity under demanding conditions.

The link between ensemble and exemplar coding observed in the present study is consistent with a recent individual differences study which examined the relationship between exemplar and ensemble coding for different high- and low-level properties, including face identity. Using a direct assessment of ensemble coding, whereby participants determined the mean identity of a set of four faces, the authors reported a very strong positive correlation, $r(45) = .76$, between the coding of face identity exemplars and ensembles (Haberman et al., 2015). They concluded that there is a mutual information transfer between ensemble and individual representations, compatible with shared limits on individual and ensemble processing. Our results are compatible with this interpretation, by providing direct evidence for a reduction in ensemble coding when participants were given less opportunity to encode individual identities. Note that the link between exemplar and ensemble identity coding observed in both the present and previous data does not imply any specific directionality of information transfer (Haberman et al., 2015), and thus cannot be taken as evidence that ensemble identity coding requires exemplars, although this is one possibility. The reverse direction is also possible, with ensemble identity driving the coding of exemplars (Hochstein et al., 2015). Finally, ensemble and exemplar coding of identity could be based on independent processes, and the observed reduction in representation strength at short durations and for large set sizes in the present study could be because both processes are subject to similar capacity restrictions. Although similar capacity limits for ensemble and exemplar coding appears to be inconsistent with the assumption that a main benefit of computing ensemble representations is to overcome the severe capacity limitations of our visual system (Alvarez, 2011), there is nevertheless evidence for partial independence of ensemble and exemplar coding for identity. For instance, a recent study demonstrated that ensemble and identity information develop dissociably across childhood and adolecence (Rhodes et al., in press), suggesting that separable processes might underlie ensemble and exemplar coding of identity for groups of faces.

On a related note, we cannot determine whether participants made their responses at short durations based on their coding of ensembles, exemplars, or a combination of both. For instance, participants could endorse a matching set average not because it matches the corresponding ensemble identity representation, but because it is more similar to the matching exemplar representations than a mismatching set average is. Thus, without the finding of the “classic” dissociation of preserved ensemble coding when exemplar coding is inaccurate or absent, it is not possible to determine which of these processes has actually taken place and driven participants’ responses. However, the present data show for the first time that – whichever process(es) drives responses at short durations – ensemble coding of identity does not provide a “short-cut” to encoding the identity of a group of faces within very brief durations, indicating a limitation in the power of averaging that in the past has been demonstrated for a wide range of other types of information.

Another recent study has examined the relationship between ensemble and exemplar coding but for facial expressions rather than identity (Li et al., 2016). In their first experiment, Li and colleagues used a membership-identification paradigm and also varied presentation duration. An association between individual and ensemble processing was observed, whereby the tendency to endorse both the mean and individual member expressions decreased with reduced exposure times. However, despite the descreased strength, ensemble coding was still significantly stronger than exemplar coding at shortest durations of 50 ms, 500 ms, and 1500 ms, suggesting that dissociable processes are underlying ensemble and exemplar coding. The authors concluded that ensemble and exemplar coding processes are separate processes that interact with each other, such that an increase in the precision of individual representations, for instance as a result of more time for encoding, also increases precision of the ensemble representation for facial expressions. Our study suggests that there might be an even stronger link between ensemble and exemplar coding for face identity than for expression, as we found no evidence for a dissociation between these processes at short durations or for large set sizes. This stronger link could suggest that an ensemble coding process is less efficient for face identity than for expression, or that separate processes might underly ensemble coding for these different properties.

In their individual differences study, Haberman and colleagues (2015) demonstrated the existence of separate, potentially domain-specific, ensemble processes. Specifically, they found evidence for at least two separate ensemble coding process, one designated for processing of low-level features (e.g., orientation and color of geometric shapes) and one for processing of high-level features (e.g., expression and identity of faces). While there were generally weak correlations between ensemble tasks across feature domains (rs < .30, n.s.), performance on different ensemble tasks within the low-level feature domain (e.g., orientation vs. colour) were strongly correlated (rs > .50, ps < .05), indicating that low-level ensemble perception might rely on a single ensemble coding process. However, there was a weaker relationship ($r = .42$, $p = .10$) within the high-level feature domain (expression vs. identity of faces), suggesting that ensemble coding for expression and identity of faces could be driven by partially distinct processes. Our data, combined with those of Li et al. (2016), suggest that rapid ensemble coding differs for identity and expression. Differences in ensemble coding of expression and identity may not be surprising, given that identity is an invariant characeristic of a person, whereas expressions are highly variable and may change from one moment to another. Representations for invariant aspects of faces are distinct to representations for changeable aspects (Haxby, Hoffman, & Gobbini, 2000), and it is possible that distinct ensemble coding processes are involved in coding invariant and changeable facial attributes.

Our data also provided some support for the idea that ensemble and exemplar coding can be partially dissociated. When groups were presented for 3200 ms or longer, ensemble identity representations were significantly reduced compared to those of individual exemplars. One interpretation of this result is that ensemble coding of identity might reach a maximum when (small) groups of four faces are seen for between 800 and 3200 ms, and then decay more rapidly than exemplar representations, which continue to build up. A slightly different interpretation would be that ensemble identity coding might cease, or might not be utilized although being coded, when participants are given enough time to encode each face. The dissociation between ensemble and exemplar coding at longer presentation durations is consistent with previous observations for non-face stimuli, e.g., in a study of perceived numerosity (Corbett, Oriet, & Rensink, 2006). When given unlimited time to inspect sets of digits, and asked to determine which of two sets had the larger average value, participants utilized an exemplar-based “counting” strategy. In contrast, when they had restricted brief viewing times between 80 and 650 ms they used ensemble coding. Similarly, ensemble size information for sets of circles occurred reliably across all observers when presentation duration was short (200 ms), but was no longer obligatory for some participants when they were given 1000 ms time to encode the set (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014). Together with these previous findings, the results of the present study suggest that ensemble coding may no longer be utilized when encoding conditions allow detailed coding of exemplar information.

If ensemble coding of facial identities is linked to the coding of individual face identity, and the individual exemplars are retained, as shown here and elsewhere (Neumann et al., 2013), then the additional benefit for coding the ensemble identity representation is not clear. So far, ensemble coding has been shown to be important for scene perception (Green & Oliva, 2009), guidance of attention to outliers (Brady & Tenenbaum, 2013), and guidance of visual search (Corbett & Melcher, 2014). One specific function of ensemble coding of face identity could be that it contributes to the norm-based coding of face identity. The abstraction of average properties plays an important role for coding of perceptual norms, which may function as references for representations of individual face identities (Leopold, O'Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006). One interesting avenue for further research, which we are investigating, is whether ensemble coding is related to norm-based coding.

Three additional points are worth noting about the set size manipulation. First, the experiment revealed a tight relationship between exemplar and ensemble coding, with a strong reduction in both representation strengths as set size increased. The method we used to manipulate set size, by adding new *different* faces (and thus making the group more heterogeneous), may have contributed to the observed reduction in ensemble coding. Previous studies that found ensemble coding for large sets typically increased set size by repeating some, or all, of the set items (e.g., Ariely, 2001; Haberman & Whitney, 2009). We decided against repeating face identities, because this would have resulted in highly unrealistic groups that could not plausibly occur in real life. However, the heterogeneity of a set can affect ensemble coding. For instance, precision estimating the mean sizes of sets of circles was reduced when set size was increased by adding *new* circle exemplars, such that the group consisted of circles of unique sizes (Marchant, Simons, & de Fockert, 2013). In contrast, no significant change in precision was found when set size was increased by simply adding circles of sizes that already existed in the set. Thus, ensemble representations might be coded for larger groups, but only if these contain a limited number of *different* exemplars. The use of different identities here may thus have contributed to a lack of ensemble coding for large sets. Importantly, however, our results suggest that ensemble coding of face identity incorporates only a limited number of different faces.

Second, set size manipulations have previously been used to determine whether or not ensemble coding actually incorporates information about all exemplars of a presented set, or whether it is capacity-limited, such that only a selection of exemplars is sampled. Several studies have reported evidence compatible with capacity limits for ensemble coding of circle size (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Gorea, Belkoura, & Solomon, 2014; Marchant et al., 2013), and the number of sampled exemplars has been estimated at about four individual circles (Allik et al., 2013; Gorea et al., 2014). The present data indicate that capacity limits may also exist for ensemble coding of facial identity. In Experiment 1, a strong reduction of endorsement scores was already observed between set sizes “2” and “4”. Thus, the capacity for ensemble coding of face identity appears to be more limited than that for coding size. This should not be surprising, given the greater complexity of faces, and the strong attentional capacity limits on the coding of *individual* face identities (Bindemann, Burton, & Jenkins, 2005; Neumann & Schweinberger, 2009; Palermo & Rhodes, 2002).

Third, a potential limitation of Experiment 1 is that the set average probe faces might look increasingly different from exemplar faces (e.g., due to increases in smoothness of skin texture) as more faces are included in the averages. Thus the reduction in ensemble coding with increasing set size could simply reflect this change in image quality. We argue that this is not the case, however, because any increase in rejection rates for set average probes based on changes in image quality should occur irrespective of whether the probes are “matches” or “mismatches” for the previous set and this was not the case. Inspection of Table 1 shows a reduction in “present” responses with increasing set size only for “matching” set averages, whereas for “mismatching” set averages “present” responses actually became *more* frequent when set size was increased from 2 to 4, and remained stable across larger set sizes. Thus, the observed reduction in endorsement scores for set averages with increased set size cannot simply be explained by a strategy to reject set averages based on image quality differences[2].

In summary, the present study provides no evidence for ensemble coding occurring when exemplar identification failed as a result of either large set size (Experiment 1) or brief presentation durations (Experiment 2). Instead, increasing set size and reducing presentation duration reduced exemplar and ensemble coding strength to a remarkably similar extent. Moreover, ensemble coding was only found for conditions that allowed the coding of exemplar representations for the faces of the group. This property of ensemble face identity coding is an important distinction to ensemble coding of other information, e.g., the set size of circles, which is often found even though individual items were not coded accurately. Given the importance of distinguishing individuals for social interaction, coding the *individual* identity of each face might be the default way to process faces, even when they are in a group. Ensemble coding strength was generally tightly coupled with exemplar coding strength, with one exception. That is, ensemble coding of face identity was reduced at longer durations that enhanced exemplar coding of identity. Such a dissociation suggests that ensemble information might not be required when exemplar information can be accurately coded. It remains to be seen whether this is a particular quality of face identity or whether this is also the case for other complex stimuli.

[2] It should also be emphasized that great care was taken to minimize the risk of artifacts in the morphed set average faces (e.g., using full frontal faces of neutral expression, standard mask that excluded facial hair). Visual inspection suggested only subtle changes in image quality between averages of two and more than two faces, and no obvious changes between set averages for set sizes 4, 6, and 8. However, given the strong endorsements of set averages at set size 2, it may be that the averages in this condition, composed of only two faces, may resemble the targets more so than in conditions with more faces.

References

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. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122-131.

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

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.

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition, 98*(2), 177-197.

Brady, T. F., & Tenenbaum, J. B. (2013). A Probabilistic Model of Visual Working Memory: Incorporating Higher Order Regularities Into Working Memory Capacity Estimates. *Psychological Review, 120*(1), 85-109.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1), 286-291.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 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*(7), 891-900.

Corbett, J. E., & Melcher, D. (2014). Stable Statistical Representations Facilitate Visual Search. *Journal of Experimental Psychology-Human Perception and Performance, 40*(5), 1915-1925.

Corbett, J. E., Oriet, C., & Rensink, R. A. (2006). The rapid extraction of numeric meaning. *Vision Research, 46*(10), 1559-1573.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181-3192.

de Fockert, J., & Gautrey, B. (2013). Greater visual averaging of face identity for own-gender faces. *Psychonomic Bulletin & Review, 20*(3), 468-473.

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

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of vision, 14*(9).

Green, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*, 137-176.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual Differences in Ensemble Perception Reveal Multiple, Independent Levels of Ensemble Representation. *Journal of Experimental Psychology-General, 144*(2), 432-446.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets 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.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception & Psychophysics, 72*(7), 1825-1838.

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review, 18*(5), 855-859.
Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.
Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision, 15*(4), 7-7.
Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1571), 1671-1683.
Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision, 15*(4), 9.
Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience, 4*(1), 89-94.
Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., et al. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in Psychology, 7*(1332).
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.
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*(2), 245-250.
Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods Instruments & Computers, 36*(4), 630-633.
Neumann, M. F., & Schweinberger, S. R. (2009). N250r ERP repetition effects from distractor faces when attending to another face under load: Evidence for a face attention resource. *Brain Research, 1270*, 64-77.
Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers abstract mean and individual information from sets of famous faces. *Cognition*(128), 56-63.
O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., et al. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research, 51*(1), 74-83.
Oliva, A., & Torralba, A. (2006). *Building the gist of a scene: the role of global image features in recognition* (Vol. 155). Amsterdam: Elsevier Science BV.
Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82*(3), 225-257.
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.
Pavlovskaya, M., Soroker, N., Bonneh, Y. S., & Hochstein, S. (2015). Computing an Average When Part of the Population Is Not Perceived. *Journal of Cognitive Neuroscience, 27*(7), 1397-1411.
Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans Use Summary Statistics to Perceive Auditory Sequences. *Psychological Science*.
Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46*(18), 2977-2987.
Rhodes, G., Neumann, M. F., Ewing, L., Bank, S., Read, A., Engfors, L. M., et al. (in press). Ensemble coding of faces occurs in children and develops dissociably from coding of individual faces. *Developmental Science*.
Rhodes, G., Neumann, M. F., Ewing, L., & Palermo, R. (2015). Reduced set averaging of face identity in children and adolescents with autism. *Quarterly Journal of Experimental Psychology, 68*(7), 1391-1403.
Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11*(12), 8.

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*, *25*(10), 1903-1913.

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays - The integration of direction information. *Vision Research*, *29*(1), 47-59.

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From Textures to Crowds: Multiple Levels of Summary Statistical Perception. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 695-710). Cambridge, Massachusetts: The MIT Press.