Cognition 128 (2013) 56–63

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

# Viewers extract mean and individual identity from sets of famous faces

Markus F. Neumann [a,*], Stefan R. Schweinberger [a,b], A. Mike Burton [c]

[a] Department of General Psychology, University of Jena, Germany
[b] DFG Research Unit Person Perception, University of Jena, Germany
[c] School of Psychology, University of Aberdeen, UK

ARTICLE INFO

*Article history:*
Received 12 April 2012
Revised 18 January 2013
Accepted 14 March 2013
Available online 13 April 2013

*Keywords:*
Set representation
Ensemble coding
Face
Identity
Averaging

ABSTRACT

When viewers are shown sets of similar objects (for example circles), they may extract summary information (e.g., average size) while retaining almost no information about the individual items. A similar observation can be made when using sets of unfamiliar faces: Viewers tend to merge identity or expression information from the set exemplars into a single abstract representation, the set average. Here, across four experiments, sets of well-known, famous faces were presented. In response to a subsequent probe, viewers recognized the individual faces very accurately. However, they also reported having seen a merged 'average' of these faces. These findings suggest abstraction of set characteristics even in circumstances which favor individuation of the items. Moreover, the present data suggest that, although seemingly incompatible, exemplar and average representations co-exist for sets consisting of famous faces. This result suggests that representations are simultaneously formed at multiple levels of abstraction.



## 1. Introduction

"Set representations" have recently attracted increasing research interest. When seeing groups of perceptually similar objects, information such as size, or motion, may be coded via summary statistics in terms of a mean value across exemplars (Albrecht & Scholl, 2010; Chong & Treisman, 2003). Whenever observers can capitalize on redundancy of information – a common observation in structured sets – they can compress this information into a single representation such as the set average (Alvarez, 2011). In a seminal investigation, Ariely (2001) investigated size representations from sets containing differently sized circles. Critically, participants tended to identify a test circle as having been presented when it had a similar size to the mean of the whole set, even when such an item had not been present. Moreover, participants were near chance when they had to choose which of two circles had been presented. Taken together, these findings suggest that (i) mean size information was computed and retained for the set and (ii) size information of individual set members was unavailable. There are different potential explanations for weak exemplar representations. First, encoding of precise exemplar representations may not routinely occur, or may simply contain too much noise, perhaps due to the lack of focal attention to set exemplars. Alternatively, an individual representation may initially be computed but may then be discarded extremely fast.

Recently, statistical representations have been demonstrated for sets of perceptually complex stimuli, such as faces. When asked to compare the emotional intensity of a single image with the mean of a set (up to 16 face photographs varying in emotional intensity), participants performed highly accurately (Haberman & Whitney, 2007, 2009). Performance was actually comparable to a control "exemplar" condition, in which participants compared an image with a homogeneous set with constant emotional

* Corresponding author. Address: School of Psychology, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. Tel.: +61 8 6488 1131; fax: +61 8 6488 1006.

*E-mail address:* markus.neumann@uwa.edu.au (M.F. Neumann).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.03.006

intensity. Beyond extraction of mean emotion (and gender, see Haberman & Whitney, 2007) information from sets of faces, a similar mechanism may compute the mean *identity* from sets of *unfamiliar* faces. In one recent study (de Fockert & Wolfenstein, 2009), participants initially saw sets containing photographs of 4 unfamiliar faces from different individuals. In a "match" condition, a subsequent single image could either be an exemplar image from the previous set, or an average morph created from the four set images. Strikingly, the set averages (which had never been seen) received more "present" responses than the (seen) exemplars. The authors concluded that averaging identity information might serve as the "default mode" for generating mental representations from groups of faces.

Given that facial representations should serve person recognition, this is a surprising finding, since mean identity representations should actually *prevent* identification of a specific person in a group. It is relatively straightforward to understand how superficial averaging of abstract shapes might take place in the visual system, but much harder to account for averaging over such high-level characteristics as someone's identity. For this reason, it is important to note that the authors used unfamiliar faces. Crucially, unfamiliar face recognition is strongly image-dependent and sensitive to superficial picture similarity (Bruce et al., 1999), and is thus based on very different mechanisms than familiar face recognition. For example, viewers are very good at matching different images of a familiar person, but very poor at matching unfamiliar faces (Bruce, Henderson, Newman, & Burton, 2001; Burton, Bruce, & Hancock, 1999; Clutterbuck & Johnston, 2004; Kemp, Towell, & Pike, 1997). This discrepancy suggests a qualitative difference in perception of familiar and unfamiliar face identities (Hancock, Bruce, & Burton, 2000), which may also have consequences for the interpretation of the identity set averaging data. Accordingly, increased percentages of "present" responses to matching averages in the study of de Fockert and Wolfenstein (2009) could reflect *image* averaging across similar pictures, rather than *identity* averaging. If viewers are failing to differentiate between the unfamiliar people shown to them, they might plausibly construct a set average combining these images. So, while this study certainly demonstrates set averaging for a class of high-level stimuli (faces), we argue that evidence for *identity* set averaging would be much more compelling if it could also be shown to exist for familiar faces sets.

Another important characteristic of previous studies examining set averaging for faces was relatively small image variability within sets. For instance, set averaging for facial expressions was generally investigated by assembling sets from a single identity, using slightly different emotional intensities from a morph continuum between two veridical expressions (Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007, 2010). One study on set *identity* averaging actually involved 4 true set photographs, but had sets deliberately arranged to comprise similar identities (de Fockert & Wolfenstein, 2009). Therefore, low recognition rates for set exemplars may have originated from participants being unable to differentiate between exemplars at encoding. It is important to see if the use of more naturally diverse sets could increase exemplar memory, and whether this would in turn affect the quality and strength of set representations.

In sum, previous studies have investigated set averaging using face sets that varied little on either identity or image properties. In the present study, we tested facial identity averaging by using diverse pictures from highly familiar identities, for which participants have rich pre-existing mental representations. We further encouraged identity processing for half of the participants by instructing them to indicate whether a specific *person* had been seen in a set of faces, while the other half indicated whether a specific *image* had occurred. We expected that set averaging would be strongly reduced or absent for highly familiar faces, and that performance would reflect accurate representation of exemplars instead; since viewers know these identities, and faces in the set were quite diverse, there appears to be no advantage in averaging across them.

## 2. Material and methods

The present article includes four experiments that share the following aspects. Stimuli were 240 original faces collected from various internet sources, 10 each from 24 well-known celebrities (12 German and 12 International). Sixty gender-homogeneous sets were created from these photographs, each containing 4 images of different identities. Images contributing to a set were chosen to be roughly similar with respect to head angle and gaze direction. Five sets from 12 different identity combinations were assembled. Note that as a result of obtaining the images from the internet, image variation within the sets was large. All images were taken under entirely non-standardized conditions, causing considerable variation on image parameters such as lighting. Additional *set averages* were created for each of the 60 sets by morphing across the respective 4 set images. Image size was $247 \times 387$ pixels, all images were presented gray-scaled and fitted in an oval mask, excluding most of the hair.

Set displays contained 4 images randomly assigned to 4 specified positions on the screen (cf. Fig. 1), and were presented for 1500 ms. Immediately following the set display (ISI = 0), probe images were displayed for 500 ms, in smaller size than the set images ($200 \times 300$ pixels). Participants used both index fingers to indicate via button press ("f" and "j" on a standard German keyboard) whether or not the probe image had been present in the previous S1 set. Probe images were: (i) a set exemplar (i.e., an image from the previous set); (ii) a new exemplar of one of the 4 identities of the previous set; (iii) a new exemplar of a different familiar identity; (iv) the average of the 4 set images; (v) the average of 4 different images of the set identities; or (vi) the average of 4 images of different familiar identities.

In each of these six conditions, 60 trials were presented, with 10 trials per condition in each of 6 experimental blocks. Response button assignment for "present" and "absent" was counterbalanced across participants. A blank screen for 2200 ms allowed for a total response window of 2700 ms.