

# Stimulus set size modulates the sex–emotion interaction in face categorization

Ottmar V. Lipp · Fika Karnadewi · Belinda M. Craig · Sophie L. Cronin

Published online: 5 March 2015  
© The Psychonomic Society, Inc. 2015

**Abstract** Previous research has shown that invariant facial features—for example, sex—and variant facial features—for example, emotional expressions—interact during face categorization. The nature of this interaction is a matter of dispute, however, and has been reported as either asymmetrical, such that sex cues influence emotion perception but emotional expressions do not affect the perception of sex, or symmetrical, such that sex and emotion cues each reciprocally influence the categorization of the other. In the present research, we identified stimulus set size as the critical factor leading to this disparity. Using faces drawn from different databases, in two separate experiments we replicated the finding of a symmetrical interaction between face sex and emotional expression when larger sets of posers were used. Using a subset of four posers, in the same setups, however, did not provide evidence for a symmetrical interaction, which is also consistent with prior research. This pattern of results suggests that different strategies may be used to categorize aspects of faces that are encountered repeatedly.

**Keywords** Face categorization · Sex · Emotional expression · Stimulus set size

The human face is a uniquely rich source of information that is essential for our functioning in a social world. Facial cues like eye gaze and emotional expressions can give us insight into what others may be thinking or feeling, as well as whether to

approach or avoid them. Additionally, faces signal a person's gender, age, ethnicity, and identity. Early research was directed at understanding the processing of single facial characteristics in isolation, uncovering findings such as the “happy face advantage,” the faster categorization of happy than of angry faces (Leppänen, Tenhunen, & Hietanen, 2004), as well as the faster categorization of other-age or other-race faces than of same-age or same-race faces (Ge et al., 2009; Zhao & Bentin, 2008). More recently, the interaction of multiple cues such as sex and emotion is becoming better understood. Investigation of these interactions is important, because they can dissociate theories of face processing that either propose independent routes for the processing of sex and emotion cues (Bruce & Young, 1986; Young & Bruce, 2011) or point to the idea that processing of these facial features shares common neural substrates (Haxby, Hoffman, & Gobbini, 2000).

Studies investigating the interaction between cues of sex and emotional expression have yielded reliable results, suggesting that cues of femininity are associated with happiness, whereas masculinity is associated with anger (see Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007; Hess, Thibault, Adams, & Kleck, 2010). This conclusion is based on numerous lines of evidence including rating studies and self-report data (e.g., Hess, Adams, & Kleck, 2004; Hess & Bourgeois, 2010), as well as reaction time based categorization tasks. Using such a categorization task, Aguado, Garcia-Gutierrez, and Serrano-Pedraza (2009) provided a clear demonstration of a symmetrical interaction between poser sex and emotional expression information. Participants were presented with facial images of 32 individuals, half male and half female, expressing either happiness or anger. These faces were presented in two categorization tasks, one requiring a sex judgment and the other requiring an emotional expression judgment. Participants were faster to categorize the sex of happy than of angry females and faster to categorize the expression of angry males than of angry females. Moreover, this

---

O. V. Lipp (✉)  
School of Psychology and Speech Pathology, Curtin University,  
Bentley, Western Australia 6845, Australia  
e-mail: Ottmar.Lipp@curtin.edu.au

F. Karnadewi · B. M. Craig · S. L. Cronin  
School of Psychology, University of Queensland,  
St. Lucia, Queensland, 4072 Australia

interaction was evident also when only the top or bottom segments of the faces were presented or when the faces were inverted (Aguado et al., 2009, Exps. 2 and 3).

These findings suggest a symmetrical interaction between sex and emotion categorizations. However, using the Garner paradigm, Le Gal and Bruce (2002) failed to find an interaction between sex and emotion cues. Atkinson, Tipples, Burt, and Young (2005) and Karnadewi and Lipp (2011) found evidence for an asymmetrical interaction. In the Garner paradigm (Garner, 1974, 1976, 1983), categorization performance in an *orthogonal condition* is compared with performance in a *control condition*. In the control condition, the stimuli vary only on the task-relevant dimension (e.g., participants categorize happy male and female faces by sex), whereas in the orthogonal condition, stimuli also vary on a task-irrelevant dimension (e.g., participants categorize angry and happy male and female faces by sex). Slower performance in the orthogonal than in the control condition indicates that variation on the task-irrelevant dimension interferes with the processing of the task-relevant dimension. Using this approach, Atkinson et al. (2005) and Karnadewi and Lipp (2011) were able to show that variation in poser sex slowed the categorization of emotional expressions, whereas variations in emotional expressions did not slow sex categorization.

It could be argued that these disparate results are due to differences in the ways that the effects of task-irrelevant dimensions are assessed—within a single orthogonal task (Aguado et al., 2009) or between orthogonal and control tasks (Atkinson et al., 2005; Karnadewi & Lipp, 2011). This, however, seems unlikely. Atkinson et al. analyzed the data from their orthogonal conditions in  $2 \times 2$  (Sex  $\times$  Expression) factorial analyses of variance (ANOVAs) and failed to find evidence for a symmetrical interaction. Similar analyses of our data (Karnadewi & Lipp, 2011) yielded equally nonsignificant outcomes, with all  $F$  values  $< 2.0$ ,  $p > .20$ . So what drives the differential results obtained in the “orthogonal tasks” used by Aguado et al. and by Atkinson et al. and Karnadewi and Lipp? The experimental procedures differed in a number of details, such as the emotional expressions used (Atkinson used fearful and happy; Aguado et al. and Karnadewi & Lipp used angry and happy), whether the type of judgment was between (Atkinson et al., 2005; Karnadewi & Lipp, 2011) or within (Aguado et al., 2009) subjects, or the database used to obtain the stimuli (Atkinson et al., 2005: caricatured versions of the Ekman & Friesen, 1976, faces; Aguado et al., 2009: Ekman & Friesen, 1976, and the Karolinska Directed Emotional Faces database, Lundqvist & Litton, 1998; Karnadewi & Lipp: NimStim faces, Tottenham et al., 2009). The latter difference may be important, since the faces differed in the extent of teeth display, which has been shown to affect performance in visual search for emotional expressions (Horstmann, Lipp, & Becker, 2012). Additionally, there was a difference in the number of individuals shown across studies. Atkinson et al.

and Karnadewi and Lipp employed pictures of one male and one female poser, each showing a happy and an angry expression. So, although four different pictures were presented, only two individuals were shown to each participant. Aguado et al., on the other hand, employed pictures of 32 different individuals: eight happy males, eight happy females, eight angry males, and eight angry females. Thus, whereas participants in the studies by Atkinson et al. and Karnadewi and Lipp saw the same pictures repeatedly, no individual was repeated in the studies by Aguado et al. (who also discussed this difference across studies).

Prior research has shown that the processing of faces changes across repeated presentations. Hart, Whalen, Shin, McNerney, Fischer, and Rauch (2000) demonstrated that racial in-group and out-group faces elicited the same extent of amygdala activation when seen first, but that amygdala activation by racial in-group faces decreased during a second encounter whereas amygdala activation in response to racial out-group faces remained high. A similar finding emerged in a gender categorization task, in which increased activation in the left amygdala was observed in response to novel faces whereas a relative decrease of activation in early visual areas was apparent for faces that had been encountered before (Dubois et al., 1999). This implies that novel and repeatedly presented faces are processed differently, with the changes in activation of early visual areas possibly reflecting an increased ease of processing the sex cues following repeated presentation. Consistent with this interpretation, Quinn and Macrae (2005) reported shorter sex categorization times for faces that had been repeated, but only if the prior exposure also required an explicit sex categorization. Relatedly, Craig, Mallan, and Lipp (2012) demonstrated that the influence of another invariant cue, race, on emotion categorization depended, amongst other things, on the stimulus set size.

The present study was designed to assess whether the nature of the interaction between face sex and emotional expressions in face categorization is affected by stimulus set size. On the basis of the results reported by Aguado et al. (2009), we expected to replicate the symmetrical interaction between face sex and emotional expression when pictures of a large number of different individuals were presented. We expected this effect to be reduced or absent if pictures of only a small number of individuals were presented repeatedly.

### Experiment 1a: Categorization of 32 individuals

The purpose of Experiment 1a was to replicate the results reported by Aguado et al. (2009). Participants were presented with the same set of 32 pictures used by Aguado et al. and were asked to perform a sex and an expression categorization task.

## Method

**Participants** A total of 24 volunteers (four males, 20 females) received course credit in return for their participation. All of the participants (age range: 17–37 years,  $M = 19.50$  years) had normal or corrected-to-normal vision. Participant numbers were based on those in previous studies (Aguado et al., 2009; Atkinson et al., 2005; Karnadewi & Lipp, 2011), which had used samples of 16 per task condition.

**Stimulus materials** Thirty-two grayscale pictures were used; half depicted angry and half depicted happy faces. All of the images were of different individuals with their hair removed and teeth displayed in the happy images. The posers were the same as those used by Aguado et al. (2009). The stimuli were compiled from the Karolinska Directed Emotional Faces (KDEF; Lundqvist & Litton, 1998) and the Pictures of Facial Affect (POFA; Ekman & Friesen, 1976) databases. The stimulus set comprised pictures of eight happy males (KDEF: AM07, AM22, BM29, BM12, AM13, AM11, AM23, AM16), eight happy females (KDEF: AF19, BF04, AF11, AF25, AF02, AF01, AF27, AF28), eight angry males (KDEF: AM03, BM17, AM05, AM09, AM10, BM15, BM21; POFA: IMG0106), and eight angry females (KDEF: AF07, BF26, AF14, AF35, AF31, BF161 POFA: IMG0089, IMG0096). The faces were set to grayscale, edited to a size of  $220 \times 250$  pixels, and dropped in a gray background.

**Apparatus** Stimulus presentation was controlled by DMDX (Forster & Forster, 2003) on a 17-in. CRT monitor (resolution:  $1,280 \times 1,024$  pixels, with 75-Hz refresh rate). Participants were provided with a button box—equipped with two buttons that were labeled in accordance with the task (sex categorization, “male” vs. “female”; or expression categorization, “happy” vs. “angry”)—on which to record their responses. The button labeling (e.g., “angry” on the left or right button, etc.) was counterbalanced across participants.

**Procedure** Each participant was tested individually and completed four face categorization tasks, two of which are not pertinent to the present report. In these two unrelated tasks, participants were presented with a subset of the faces used in the tasks reported here. Preliminary analyses showed no difference in task performance when the tasks reported here were performed either first or second. After providing informed consent, participants were seated at an approximate viewing distance of 40 cm from the monitor. They were then told that they would perform several tasks that required them to categorize a face on either its expression or sex, as quickly as possible, but without sacrificing accuracy. Prior to the experimental trials, a written version of the instructions appeared on the screen.

Participants completed both a sex and an expression categorization task, in counterbalanced order. In each task, the

faces varied in sex and emotion cues, and pictures of 32 different individuals were presented. Deviating from the original procedure, participants completed each task twice, to yield a total of 64 trials per task. Within each trial, a fixation marker (“+”) was presented in the middle of the computer screen for 500 ms and replaced by a target face. The target face was shown for 10,000 ms or until a response was made. The next trial started 2,000 ms after a response was made.

**Data reduction and analysis** Prior to the main analyses, reaction times faster than 100 ms and those that fell outside three SDs from each individual’s mean were removed. The median reaction times and error percentages were subjected to two separate  $2 \times 2$  (Sex [male vs. female]  $\times$  Expression [happy vs. angry]) repeated measures ANOVAs for the sex and expression categorization tasks. Preliminary analyses including the factor Task Sequence did not yield any interactions involving this factor, and results are reported collapsed over it.

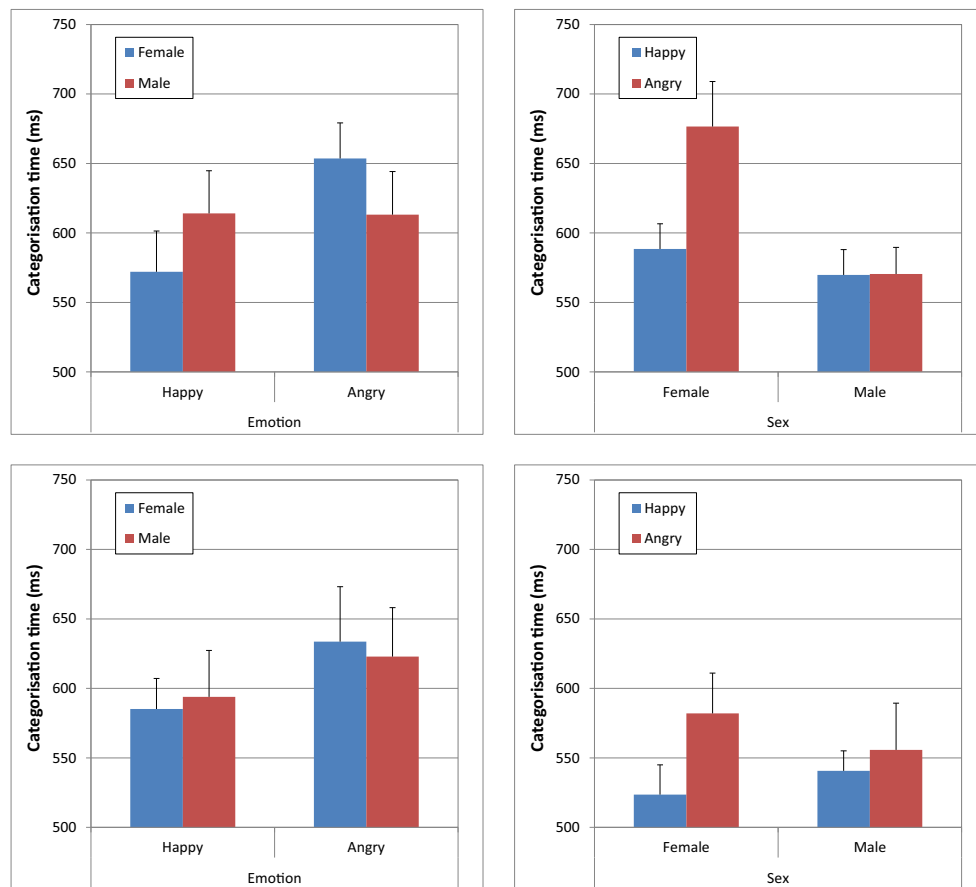
## Results

As is depicted in the upper left panel of Fig. 1, the expression task yielded a happy categorization advantage, such that happy faces were categorized faster than angry faces,  $F(1, 23) = 6.67, p = .017, \eta_p^2 = .225$ . This effect was qualified by a Sex  $\times$  Expression interaction,  $F(1, 23) = 19.73, p < .001, \eta_p^2 = .462$ . Replicating the pattern observed by Aguado et al. (2009), the happy face advantage was significant for female posers,  $t(23) = 4.53, p < .001$ , but not for male individuals,  $t < 1.00$ , n.s. Analysis of the error data yielded no significant results,  $F < 1.50$ , n.s. Error rates were low and ranged from 5.2 % to 6.5 %.

Also consistent with Aguado et al. (2009), the speed of the sex categorization varied with the posers’ emotional expressions. Significant main effects of emotional expression,  $F(1, 23) = 9.70, p = .005, \eta_p^2 = .297$ , and sex,  $F(1, 23) = 25.87, p < .001, \eta_p^2 = .529$ , were qualified by a Sex  $\times$  Expression interaction,  $F(1, 23) = 8.79, p = .007, \eta_p^2 = .276$ . As is depicted in the upper right panel of Fig. 1, angry male faces were categorized faster as male than angry female faces were categorized as female,  $t(23) = 4.65, p < .001$ . No difference was observed for the happy faces,  $t < 1.3$ , n.s. Analysis of the errors yielded main effects of expression,  $F(1, 23) = 66.87, p < .001, \eta_p^2 = .744$ , and sex,  $F(1, 23) = 110.84, p < .001, \eta_p^2 = .828$ , and a Sex  $\times$  Expression interaction,  $F(1, 23) = 62.13, p < .001, \eta_p^2 = .730$ . More errors were committed when categorizing angry female faces ( $M = 37.8\%$ ,  $SD = 17.82$ ) than in any of the other conditions (all  $M$ s  $< 6.0\%$ ).

## Discussion

The results of Experiment 1a provided a direct replication of Aguado et al. (2009) in a within-subjects design



**Fig. 1** Categorization times in the emotion (left panels) and sex categorization (right panels) tasks of Experiment 1a (upper panels) and Experiment 1b (lower panels). Error bars represent standard errors of the means

and using the same faces that had been used in the original study. A symmetrical interaction between sex and emotion was found, such that in the expression categorization task happy faces were categorized faster than angry faces when posed by females, whereas in the sex categorization task, male faces were categorized faster than female faces if they expressed anger, but not happiness. This is consistent with the standard finding that anger is more likely to be associated with males and happiness with females (Hess et al., 2004; Hugenberg & Sczesny, 2006).

Experiment 1b repeated the procedure used in Experiment 1a, but utilized a smaller number of posers within each categorization task. As we speculated in the introduction, the differential findings reported by Aguado et al. (2009) and Atkinson et al. (2005) and Karnadewi and Lipp (2011) may reflect the numbers of individuals displayed within the sex and expression tasks and the differential ease of processing sex and emotion information. On the basis of the findings reported by Le Gal and Bruce, Atkinson et al., and Karnadewi and Lipp, we predicted that there would be no symmetrical interaction between face sex and emotional expression if a smaller stimulus set of four faces were used.

## Experiment 1b: Categorization of four individuals

### Method

**Participants** Twenty-four participants (seven males, 17 females; age range: 17–28 years,  $M = 18.96$  years), who had not participated in Experiment 1a, received course credits in exchange for participation. All participants provided informed consent and had normal or corrected-to-normal vision.

**Apparatus and stimuli, procedure, and data reduction and analysis** The apparatus and stimulus materials were identical to those in Experiment 1a. All participants performed four categorization tasks, two of which formed Experiment 1b. The other two tasks used pictures of faces that were different from the ones used in the tasks comprising Experiment 1b. Preliminary analyses failed to find a difference in task performance related to whether the tasks reported here were performed first or second. The expression and sex categorization tasks involved pictures of four different individuals (i.e., angry male, happy male, angry female, happy female) that were repeated 16 times, for a total of 64 trials. Eight different sets of four pictures were created at random from the total set of 32

used in Experiment 1, and three participants each were presented with a particular set. The same faces were presented in the sex and expression categorization tasks for each participant, but across participants, all of the pictures used in Experiment 1 were employed as stimuli. Preliminary analyses confirmed that the task sequence had no effect on the pattern of results. Two 2 (Sex: male vs. female)  $\times$  2 (Expression: happy vs. angry) repeated measures ANOVAs were conducted on the data from the expression and sex categorization tasks.

## Results

As is shown in the lower left panel of Fig. 1, and in contrast to the results of Experiment 1a, the interactive pattern between sex and expression failed to emerge in the expression categorization task,  $F(1, 23) < 1.10$ . A main effect of expression was evident,  $F(1, 23) = 5.09$ ,  $p = .034$ ,  $\eta_p^2 = .181$ , indicating faster categorization of happy expressions. Error rates in the expression task were low overall,  $< 8.60\%$  across all conditions, and did not differ across conditions, all  $F < 2.50$ , n.s.

Seven participants provided insufficient response time data for the sex categorization task with 100 % errors in classifying the angry female. This confirmed the finding of Experiment 1a that some of the female posers were difficult to recognize as female. As is illustrated in the lower right panel of Fig. 1, participants were faster to categorize happy faces,  $F(1, 16) = 5.45$ ,  $p = .033$ ,  $\eta_p^2 = .25$ , but no interaction between expression and poser sex was obtained in the sex categorization task,  $F < 1.40$ , n.s. The analysis of the error data yielded no significant results, with all error rates being below 7 %, all  $F < 1.0$ , n.s.

## Discussion

Experiment 1b failed to replicate the pattern of results reported by Aguado et al. (2009), in that no interaction between poser sex and emotional expression was found if participants were presented repeatedly with a subset of the faces used by Aguado et al. and in Experiment 1a. This evidence for independent processing of sex and emotion cues is consistent with the results reported by Le Gal and Bruce (2002). However, it also revealed that our participants found it difficult to correctly categorize the sex of some of the angry female faces. This finding is consistent with the data reported by Aguado et al. (error rate of 26 % for angry female,  $< 6.0\%$  for the remaining) and may reflect on the stimulus materials used. Some of the female faces were difficult to recognize as such after the hair had been removed. It should be noted, however, that this did not hold universally. Some participants correctly categorized the female angry face consistently, even though others consistently miscategorized it. Experiments 2a and 2b were designed to examine the stability of the results of Experiments 1a and

1b by using a different set of face stimuli selected from a different database. Pictures of emotionally expressive Caucasian males and females from the NimStim database (Tottenham et al., 2009) and the Montreal Set of Facial Displays of Emotion (Beaupré & Hess, 2005) were chosen.

## Experiment 2a: Categorization of 16 individuals

### Method

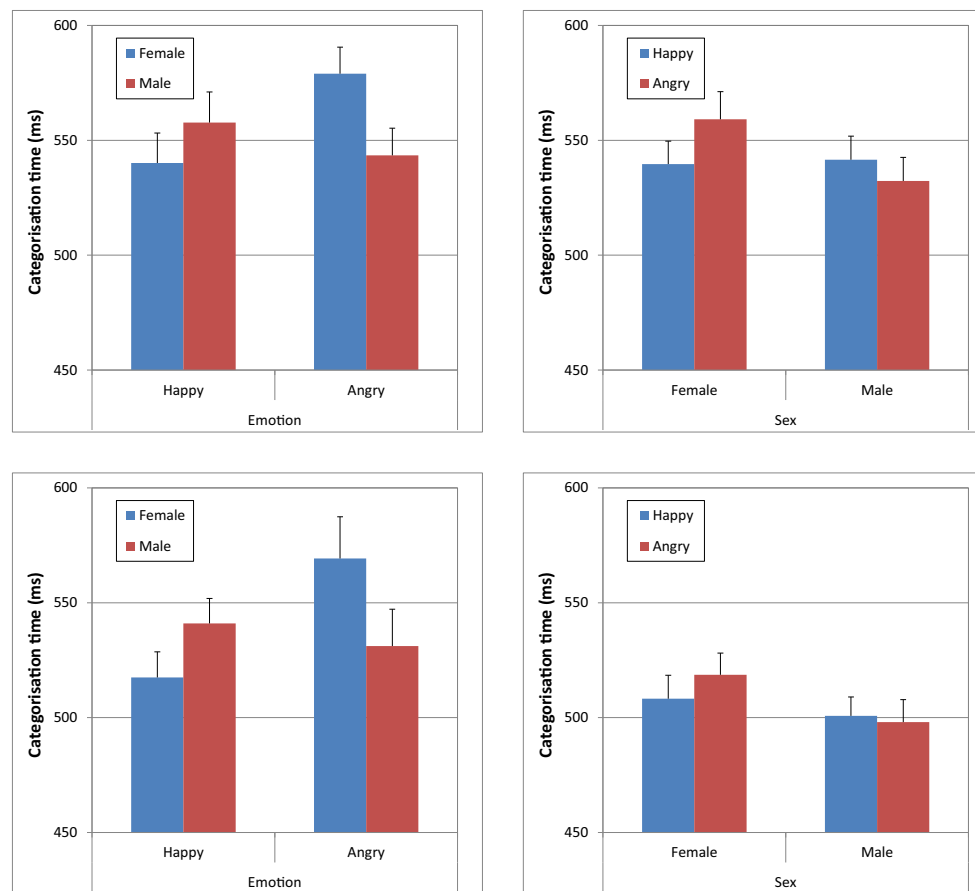
**Participants** Thirty-two naïve participants (six males, 26 females; age range: 17–32 years,  $M = 21.34$  years) received course credits in exchange for participation and were tested in groups of up to six. All participants provided informed consent and had normal or corrected-to-normal vision.

**Apparatus and stimuli, procedure, and data reduction and analysis** The images of eight male and eight female Caucasian posers, drawn from the NimStim database (Tottenham et al., 2009: Poses AN\_O and HA\_O of Models 1, 2, 3, 5, 6, 20, 21, 23, 24, and 28) and the Montreal Set of Facial Displays of Emotion (Beaupré & Hess, 2005: Poses 1 and 2 of Models 20, 22, 23, 25, 27, 28), were employed in two categorization tasks presented in counterbalanced order. Each poser provided a happy and an angry expression, resulting in 32 images of 16 different individuals. The faces were cropped of hair, set to grayscale, and dropped in a gray background at a picture size of  $187 \times 240$  pixels. The faces were presented centered in three blocks of 32 trials on a CRT monitor with a resolution of  $1,024 \times 768$  pixels and an 85-Hz refresh rate. Each face was preceded by a 500-ms fixation cross and presented for 2,000 ms or until a response was made by pressing the left or right shift key. Faces were presented in a random sequence, with the restriction that no more than four consecutive faces were of the same sex or emotion. The matches of emotion/sex to key and task order were counterbalanced across participants, and each task was preceded by eight practice trials. The stimulus displays and categorization time recording were controlled by DMDX (Forster & Forster, 2003). The data reduction and analysis were as in Experiment 1a.

### Results and discussion

As is shown in the upper left panel of Fig. 2, participants were faster to categorize pictures of males,  $F(1, 31) = 6.19$ ,  $p = .018$ ,  $\eta_p^2 = .166$ , and poser sex interacted with emotional expression in the expression categorization,  $F(1, 31) = 14.95$ ,  $p = .001$ ,  $\eta_p^2 = .325$ . Happy expressions were categorized faster than angry expressions when posed by females,  $t(31) = 3.00$ ,  $p < .001$ , whereas no such difference was observed for male posers,  $t < 1.30$ , n.s. The analysis of the error data yielded a poser Sex  $\times$  Expression interaction,  $F(1, 31) =$





**Fig. 2** Categorization times in the emotion (left panels) and sex categorization (right panels) tasks of Experiment 2a (upper panels) and Experiment 2b (lower panels). Error bars represent standard errors of the means

16.82,  $p < .001$ ,  $\eta_p^2 = .352$ , with more errors being committed when categorizing angry expressions posed by females ( $M = 10.55$ ,  $SD = 7.99$  vs.  $M = 4.43$ ,  $SD = 4.36$ ),  $t(31) = 3.67$ ,  $p = .001$ , and more errors being committed when categorizing happy expressions posed by males ( $M = 10.29$ ,  $SD = 8.60$  vs.  $M = 5.86$ ,  $SD = 4.72$ ),  $t(31) = 2.33$ ,  $p = .026$ .

The results of the sex categorization task can be seen in the upper right panel of Fig. 2. The analysis yielded a main effect for sex,  $F(1, 31) = 6.78$ ,  $p = .014$ ,  $\eta_p^2 = .180$ , and a Sex  $\times$  Expression interaction,  $F(1, 31) = 6.59$ ,  $p = .015$ ,  $\eta_p^2 = .175$ . Sex categorization was slower for angry female individuals than for angry male individuals,  $t(31) = 3.42$ ,  $p = .002$ , whereas no such difference was evident for the happy expressions,  $t < 1.0$ , n.s. Error rates were below 10 %, with the exception of angry female faces, which were misclassified on 20.18 % of the trials. The analysis yielded main effects of sex,  $F(1, 31) = 42.27$ ,  $p < .001$ ,  $\eta_p^2 = .577$ , and expression,  $F(1, 31) = 44.61$ ,  $p < .001$ ,  $\eta_p^2 = .590$ , and a Sex  $\times$  Expression interaction,  $F(1, 31) = 55.80$ ,  $p < .001$ ,  $\eta_p^2 = .643$ . Whereas there was no difference between the categorization of male,  $M = 6.25$ ,  $SD = 5.08$ , and female,  $M = 6.25$ ,  $SD = 6.44$ , happy faces,  $t < 1.0$ , n.s., more errors were committed

when categorizing female,  $M = 20.18$ ,  $SD = 7.42$ , than male,  $M = 4.56$ ,  $SD = 5.12$ , angry faces,  $t(31) = 9.58$ ,  $p < .001$ .

The results of Experiment 2a indicate a symmetrical interaction between sex and emotion in face categorization and replicate those of Experiment 1a and of Aguado et al. (2009), using a different set of faces drawn from different databases. This result emerged with a stimulus set that comprised the same number of different images used in Experiment 1a, 32, but in which only 16 different individuals were displayed. Experiment 2b was designed to assess whether the same symmetrical pattern of results would emerge if a smaller set of individuals drawn from the new face set was used. On the basis of the results reported by Le Gal and Bruce (2002), Atkinson et al. (2005), Karnadewi and Lipp (2011), and in Experiment 1b, no symmetrical interaction was expected.

### Experiment 2b: Categorization of four individuals

#### Method

**Participants** A new sample of 32 naïve volunteers (eight males, 24 females; age range: 17–24 years,  $M = 19.09$  years),

who participated in exchange for course credit, were tested in groups of up to six. All participants provided informed consent prior to the experiment and had normal or corrected-to-normal visual acuity.

*Apparatus, stimuli, procedure, and data reduction and analysis* The purpose of Experiment 2b was to confirm that no symmetrical interaction would be observed if pictures of only four individuals were viewed during expression and sex categorization. Each participant was presented with pictures of four individuals, two male and two female, each expressing happiness or anger in two categorization tasks. As in Experiment 1b, the 32 images were divided into eight separate sets, and four participants each were tested with a particular set. The apparatus used and the procedure for stimulus assignment were similar to those of Experiment 2a. The data reduction and analysis were as we described for Experiment 2a.

## Results and discussion

As is illustrated in the lower left panel of Fig. 2, expression categorization was affected by poser sex, with a happy face advantage evident for females, but not for males. The  $2 \times 2$  ANOVA confirmed this impression, yielding a main effect of sex,  $F(1, 31) = 6.19, p = .018, \eta_p^2 = .166$ , and a Sex  $\times$  Expression interaction,  $F(1, 31) = 14.95, p = .001, \eta_p^2 = .325$ . Happy faces were categorized faster than angry ones if they were posed by females,  $t(31) = 3.90, p < .001$ , but not if they were posed by males,  $t < 1.3, n.s.$  The analysis of errors yielded a Sex  $\times$  Expression interaction,  $F(1, 31) = 16.82, p = .001, \eta_p^2 = .352$ . More errors were committed when categorizing angry,  $M = 10.55, SD = 7.99$ , than happy,  $M = 4.43, SD = 4.36$ , female faces,  $t(31) = 3.66, p = .001$ , whereas more errors were committed when categorizing happy,  $M = 10.29, SD = 8.60$ , than angry,  $M = 5.86, SD = 4.72$ , male faces,  $t(31) = 2.33, p = .026$ .

Four participants failed to provide complete data, three because they miscategorized the happy male face on 100 % of the trials, and one for miscategorizing the angry female face on 100 % of the trials. As is displayed in the lower right panel of Fig. 2, participants were faster to categorize male faces,  $F(1, 27) = 6.93, p = .014, \eta_p^2 = .204$ , but sex categorization was not affected by expression [Sex  $\times$  Expression interaction:  $F(1, 27) = 1.16, p = .290, \eta_p^2 = .041$ ]. Error rates were below 13 % in all cells and were not differentially affected by expression [Sex  $\times$  Expression interaction,  $F(1, 27) = 1.95, p = .174, \eta_p^2 = .067$ ].

Resembling the pattern seen in Experiment 1, Experiment 2b failed to replicate the symmetrical interaction between face sex and expression on expression and sex categorization seen for the larger picture set used in Experiment 2a. Rather, an asymmetrical interaction was found, in that poser sex affected the categorization of emotional expressions, whereas

expression did not affect the categorization of faces by sex. This pattern deviates from that seen in Experiment 1b, in which no interaction was observed in either sex or expression categorization, but is consistent with the pattern of results reported in studies that had employed the Garner paradigm (Atkinson et al., 2005; Karnadewi & Lipp, 2011).

## General discussion

Past research has provided inconsistent findings as to the interaction of sex and expression cues in face categorization by either expression or sex. Using the Garner paradigm, Le Gal and Bruce (2002) found no evidence of an interaction, whereas Atkinson et al. (2005) and Karnadewi and Lipp (2011) found support for an asymmetrical interaction, in which sex cues affected expression categorization, but expression cues did not affect sex categorization. Aguado et al. (2009), on the other hand, found support for a symmetrical interaction, in which sex cues affected expression categorization, and vice versa. The present findings resolve this apparent contradiction by showing that stimulus set size is an important determinant of the nature of the Emotional Expression  $\times$  Sex interaction. A symmetrical interaction will emerge if the number of different images used is large—32 in the present study—but not if the number of different images used is small—four in the present study. The latter finding is also consistent with the results of post-hoc analyses reported by Atkinson et al. and performed on our data (Karnadewi & Lipp, 2011), in which no interaction emerged if the data from the Garner paradigm's orthogonal conditions were analyzed in  $2 \times 2$  factorial ANOVAs. These results resemble the results of Experiment 1b and those of Le Gal and Bruce (2002).

The present results cannot answer the question, however, of whether the size of the picture set per se or the number of different posers used is what determines whether a symmetrical interaction between poser sex and emotional expressions emerges. In Experiments 1a and 2a, we employed the same number of different images, 32, but varied the number of different posers, 32 or 16, respectively. Both experiments yielded the same pattern of results, a symmetrical interaction between poser sex and expression on face categorization. Further systematic research that keeps the number of unique images constant, but varies the number of posers used, would be required to answer this question.

Second, the present research is silent as to the mechanism that mediates the effect of stimulus set size on the interaction between face sex and expression in face categorization. One explanation might be that different strategies are used to solve the categorization tasks with large and small stimulus sets. Categorizing a large stimulus set may have required the detailed processing of each face, including both task-relevant and task-irrelevant characteristics. Thus, the task-irrelevant

characteristics—for example, sex—could impact the categorization based on the task-relevant one. For the small stimulus set, it is possible to learn the stimulus–response mapping, such that simple recognition of one of the four stimuli presented will trigger the appropriate response. One might argue that this was facilitated by using pictures of four different individuals in the tasks with small set sizes, rather than pictures of one male and one female each displaying the two expressions. Such a stimulus set was used in the orthogonal task conditions used by Atkinson et al. (2005) and Karnadewi and Lipp (2011), and this failed to yield a symmetrical interaction when analyzed in a  $2 \times 2$  factorial design. Nevertheless, Atkinson et al. and Karnadewi and Lipp reported that in their emotion tasks, performance was slower in the orthogonal condition than in the control condition, indicating that the sex information conveyed by the faces had been processed.

Consistent with this explanation, performance of the sex categorization was faster with small than with large stimulus sets in both experiments. Moreover, only in the sex tasks involving small set sizes were there cases in which particular stimuli were miscategorized on 100 % of the trials, and these miscategorizations were limited to categorizing angry females as male (eight cases) and happy males as female (three cases). Thus, the changes in facial features that lead to an angry (lowering of eyebrows, clenching of jaw muscles) or a happy (raising of eyebrows, lifting of cheeks) expression were misinterpreted as cues of masculinity or femininity, respectively (Becker et al., 2007; Hess et al., 2010). The consistency of the miscategorizations also suggests that, at least in these participants, once a particular face had been categorized as male or female, this categorization was not reconsidered, but was maintained across 16 (Exp. 1b) or 24 (Exp. 2b) presentation of the same face.

It should be noted, however, that these miscategorizations were not universal—that is, the same faces that were consistently miscategorized by some participants were categorized correctly by others. Moreover, we found no evidence for a systematic change in response patterns or speed across blocks of trials in the tasks with small stimulus sets. Post-hoc analyses of the data from these tasks that included Task Block as a factor (based on four trials for Experiment 1b and three trials in Experiment 2b, to minimize the number of participants excluded due to missing values) yielded no evidence for a symmetrical interaction during early trial blocks that disappeared as the task progressed. Such a change might be expected if one assumed that the stimulus–response mapping was learned within the initial trials of the experiment. Analyses of the data obtained during the eight practice trials completed prior to each task in Experiment 2b also failed to support such an interpretation. These analyses yielded a trend toward a Sex  $\times$  Expression interaction for the sex task where such an interaction was absent in the main task, but failed to find a similar interaction for the emotion task where such an interaction was

present in the main task. It should be noted, however, that the present research was not designed to address this question and that the results of these post-hoc analyses need to be considered with care.

The finding of an effect of face sex on expression categorization in Experiment 2b suggests that a simple strategy of learning a stimulus–response mapping is not used when it comes to categorizing faces as happy or angry. This may reflect the fact that the intensity of emotional expressions can vary across posers, and is thus less amenable to a dichotomous categorization than is face sex, which was either male or female. It may also reflect the fact that information about invariant facial characteristics is available at an earlier stage of face processing than is expression information; hence, it may be more difficult to ignore. The observation that, for small set sizes, performance in the emotion tasks was slower than performance in the sex tasks is consistent with this interpretation. This difference seems to disappear, however, if in-depth processing of each face is required, as in the tasks with large set sizes.

The present results provide evidence that the prior inconsistent reports of interactions between face sex and emotional expressions in face categorization reflect differences in the stimulus set sizes used. Symmetrical interactions are observed for large stimulus sets, but not for small sets, in which either asymmetrical interactions or no interactions emerge. The finding of an interaction between face sex and emotion in expression categorization with larger stimulus sets is of interest, since it seems inconsistent with research on the effects of another invariant facial characteristic, face race. Craig, Mallan, and Lipp (2012) followed up on studies of the effects of poser race on expression categorization that had either reported happy face advantages for same-race faces and angry face advantages for other-race faces (Hugenberg, 2005) or happy face advantages for both racial in-group and out-group faces (Kubota & Ito, 2007). Stimulus set size was, among other variables such as presentation duration and the nature of the stimuli (computer-generated or photographic), identified as one of the major determinants of the outcome. Interestingly, though, increasing the set size reduced the likelihood of finding an interaction between face race and emotional expression, a finding that is contrary to what was seen here. This may indicate that the interaction between face sex and emotional expressions is mediated by a different mechanism than is the interaction between face race and emotional expressions, and that perceptual similarity plays a larger role in the former than in the latter.

The research investigating the interaction between poser sex and emotional expressions was originally motivated by the question of whether variant face cues, such as expressions, and invariant face cues, such as sex, age, or race, are processed independently, as is proposed by Bruce and Young's model of face processing (Bruce & Young, 1986; Young & Bruce,



2011), or in a manner that makes interference likely, as is suggested in the neuroimaging-based model proposed by Haxby, Hoffman, and Gobbini (2000). Replicating previous results, the findings in the tasks with large set sizes certainly suggest an interdependence in the processing of different facial characteristics. The results of the tasks involving smaller set sizes seem to suggest a different pattern—at least for the processing of sex cues, which seems independent of variations in emotional expression. Again, this finding is consistent with the asymmetrical interactions reported in prior studies that had employed the Garner paradigm, which utilizes smaller set sizes (Atkinson et al., 2005; Karnadewi & Lipp, 2011). It is unclear as yet what drives the reduced effect of variations in emotional expressions on sex categorization for smaller set sizes. However, the finding that a symmetrical interaction between expression and sex cues is readily observable for larger face sets, in which it may be more difficult to become familiar with the individual faces, seems to argue against an account that proposes independent processing routes for variant and invariant facial cues.

In summary, the present results remove an apparent inconsistency in the literature on face categorization. They indicate that face sex and emotion will interact symmetrically if a large set of face stimuli is used in each categorization task. If the size of the stimulus set is reduced to four, an asymmetrical interaction or no interaction is observed.

**Author note** This work was supported by Grant Number DP110100460 from the Australian Research Council, to O.V.L., and a grant from UQRS/UQIRTA to F.K.

## References

- Aguado, L., Garcia-Gutierrez, A., & Serrano-Pedraza, I. (2009). Symmetrical interaction of sex and expression in face classification tasks. *Attention, Perception, & Psychophysics*, *71*, 9–25. doi:10.3758/APP.71.1.9
- Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception & Psychophysics*, *67*, 1199–1213. doi:10.3758/BF03193553
- Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, *26*, 355–370. doi:10.1177/0022022104273656
- Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*, 179–190. doi:10.1037/0022-3514.92.2.179
- Bruce, V., & Young, A. (1986). Understanding face perception. *British Journal of Psychology*, *77*, 305–327. doi:10.1111/j.2044-8295.1986.tb02199.x
- Craig, B. M., Mallan, K. M., & Lipp, O. V. (2012). The effect of poser race on the happy categorisation advantage depends on stimulus type, set size, and presentation duration. *Emotion*, *12*, 1303–1314. doi:10.1037/a0028622
- Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *NeuroImage*, *9*, 278–289. doi:10.1006/nimg.1998.0409
- Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Window display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. doi:10.3758/BF03195503
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*, 98–123. doi:10.1016/0010-0285(76)90006-2
- Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analysis* (pp. 1–38). Hillsdale, NJ: Erlbaum.
- Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., & Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*, 1199–1210. doi:10.1068/p6136
- Hart, A. J., Whalen, P. J., Shin, L. M., McInemey, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *NeuroReport*, *11*, 2351–2355. doi:10.1097/00001756-200008030-00004
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233. doi:10.1016/S1364-6613(00)01482-0
- Hess, U., Adams, R. B., Jr., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion*, *4*, 378–388. doi:10.1037/1528-3542.4.4.378
- Hess, U., & Bourgeois, P. (2010). You smile—I smile: Emotion expression in social interaction. *Biological Psychology*, *84*, 514–520. doi:10.1016/j.biopsycho.2009.11.001
- Hess, U., Thibault, P., Adams, R. B., Jr., & Kleck, R. E. (2010). The influence of gender, social roles, and facial appearance on perceived emotionality. *European Journal of Social Psychology*, *40*, 1310–1317. doi:10.1002/ejsp.715
- Horstmann, G., Lipp, O. V., & Becker, S. I. (2012). Of toothy grins and angry snarls—Open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision*, *12*(5), 7. doi:10.1167/12.5.7.1–15.
- Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion*, *5*, 267–276. doi:10.1037/1528-3542.5.3.267
- Hugenberg, K., & Sczesny, S. (2006). On wonderful women and seeing smiles: Social categorization moderates the happy face response latency advantage. *Social Cognition*, *24*, 516–539. doi:10.1521/soco.2006.24.5.516
- Karnadewi, F., & Lipp, O. V. (2011). The processing of invariant and variant face cues in the Garner Paradigm. *Emotion*, *11*, 563–571. doi:10.1037/a0021333
- Kubota, J. T., & Ito, T. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*, 738–752. doi:10.1016/j.jesp.2006.10.023
- Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Perception & Psychophysics*, *64*, 230–243. doi:10.3758/BF03195789
- Leppänen, J. M., Tenhunen, M., & Hietanen, J. K. (2004). Faster choice-reaction times to positive than to negative facial expressions: The role of cognitive and motor processes. *Journal of Psychophysiology*, *17*, 113–123. doi:10.1027/0269-8803.17.3.113
- Lundqvist, D., & Litton, J. E. (1998). *The averaged Karolinska Directed Emotional Faces—AKDEF [CD-ROM]*. Stockholm, Sweden:

- Karolinska Institutet, Department of Clinical Neuroscience, Psychology Section.
- Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology*, *88*, 467–479. doi:[10.1037/0022-3514.88.3.467](https://doi.org/10.1037/0022-3514.88.3.467)
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249. doi:[10.1016/j.psychres.2008.05.006](https://doi.org/10.1016/j.psychres.2008.05.006)
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, *102*, 959–974. doi:[10.1111/j.2044-8295.2011.02045.x](https://doi.org/10.1111/j.2044-8295.2011.02045.x)
- Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review*, *15*, 1093–1099. doi:[10.3758/PBR.15.6.1093](https://doi.org/10.3758/PBR.15.6.1093)