Vision Research 163 (2019) 14–23

Contents lists available at ScienceDirect

Vision Research

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

Gender and hometown population density interact to predict face recognition ability

Mackenzie A. Sunday[a,*], Parth A. Patel[a], Michael D. Dodd[b], Isabel Gauthier[a]

[a] *Vanderbilt University, Department of Psychology, USA*
[b] *University of Nebraska-Lincoln, Department of Psychology and Center for Brain, Biology and Behavior, USA*

A R T I C L E I N F O

Keywords:
Face recognition
Experience

A B S T R A C T

Several studies have found that individuals from small hometowns show diminished face recognition ability as compared with individuals from larger hometowns. We further this line of research by relating six measures of face recognition ability to hometown density. We predicted that the three face recognition ability measures which included a learning component would relate to hometown density whereas the three measures which did not include such a learning component would not. Instead, we found that none of the six measures related to hometown density. Interestingly, we found interactions between gender and hometown population density on many of these measures and on a general index of face recognition, with females from small hometowns outperforming males from small hometowns but no such differences in the large hometown group. In a follow-up reanalysis of a previous study, we found a similar interaction in one of two face recognition ability measures. Together, these results reveal a pattern of gender differences modulated by hometown population density. If indeed experience with faces in one's hometown influences face recognition ability, understanding these effects may require more than a quantification of the environment. Men and women growing up in the same environment likely have different experiences, which likely modulates effects on visual abilities.

1. Introduction

There is a remarkably wide range in face recognition abilities (De Bruïne, Vredeveldt, & van Koppen, 2018; Duchaine & Nakayama, 2006; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Russell, Duchaine, & Nakayama, 2009). Twin studies reveal that about half of this variability is genetic (Shakeshaft & Plomin, 2015; Wilmer et al., 2010), which leaves much of the variability unexplained. One possible additional source of variance is variability in experience with faces. Evidence for this comes from research reporting differences in face recognition ability relating to variable experience with different races (De Heering, De Liedekerke, Deboni, & Rossion, 2010; Sangrigoli, Pallier, Argenti, Ventureyra, & De Schonen, 2005; Yovel et al., 2012), different ages (Kuefner, Macchi Cassia, Picozzi, & Bricolo, 2008), and even species (Pascalis, de Haan, & Nelson, 2002). Studies regarding the cross-race effect (the tendency for individuals to more accurately recognize faces of their own race compared to faces of other races) have demonstrated the importance of childhood experience on recognition of other-race faces into adulthood (De Heering et al., 2010; Sangrigoli et al., 2005). However, mere quantity of face exposure is not sufficient to determine face recognition. Rather, this must be coupled with active categorization or discrimination among faces, emphasizing the importance of quality of face exposure (McGugin et al., 2012; Yovel et al., 2012). An other-age effect (the tendency for individuals to more accurately recognize faces of their own age in individuals who possess limited experience with other-age faces) provides further evidence that experience modulates face recognition ability (Kuefner et al., 2008). This effect varies based on compositional changes in longitudinal face exposure. For example, after five months, children who began preschool showed significantly improved recognition of child faces compared to their counterparts who were not yet in school. Both, however, showed improvement in discriminating adult faces (De Heering, Bracovic, & Maurer, 2014). Moreover, as individuals age into adolescence, improved recognition of caregiver faces (typically adult female faces) shifts to improved recognition of peer faces that share a similar pubertal status (Picci & Scherf, 2016).

Differential experience has also been proposed as an explanation for gender differences often observed in face recognition. Women often outperform men on face recognition tasks (Goldstein & Chance, 1970; Lewin & Herlitz, 2002; Lovén, Svärd, Ebner, Herlitz, & Fischer, 2013; Rehnman & Herlitz, 2007), with possible explanations for this ranging from differences in gaze preferences to differences in social motivations

* Corresponding author at: Department of Psychology, Vanderbilt University, 226 Wilson Hall, Nashville, TN 37204, USA.
E-mail address: mackenzie.a.sunday@vanderbilt.edu (M.A. Sunday).

https://doi.org/10.1016/j.visres.2019.08.006
Received 4 June 2019; Received in revised form 13 August 2019; Accepted 14 August 2019
Available online 28 August 2019
0042-6989/

(Baron-Cohen, 2002; Ingalhalikar et al., 2014; Lewin & Herlitz, 2002; Lovén et al., 2013; Sawada et al., 2014; Wolff, Kemter, Schweinberger, & Wiese, 2014; Yovel et al., 2012). One explanation is that experience could account for these differences. Ryan and Gauthier (2016) showed that performance for males and females on recognition tasks involving the faces of Barbie dolls and Transformers action figures is consistent with their reported experience. That is, females recognize Barbie faces better than males, but males recognize Transformer faces better than females, making Transformer faces the only instance of faces to date found to be recognized better by males.

Other evidence for the role experience plays in face recognition ability comes from work examining how visual deprivation affects face recognition. For example, patients deprived of early visual experience due to bilateral congenital cataracts showed deficits on the Benton Facial Recognition Test years after cataract removal, but only when head orientation and/or lighting conditions were not identical (Putzar, Hötting, & Röder, 2010). Recently, found that monkeys raised in a completely face-deprived environments exhibited different fixation patterns on faces as compared with control monkeys and also did not develop face-specific cortical patches. This suggests that early experience may be especially influential.

Significant visual deprivation, though, is not necessary to demonstrate the long-term effects of visual experience on perception. Balas and Saville (2015; 2017) demonstrated an effect of hometown population size on performance on the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006). Specifically, they found that individuals from small hometowns (less than 1000 people) showed poorer face learning ability than individuals from large hometowns (30,000–100,000 people). Participants from small hometowns also had a N170 neural response that was less specific to faces over objects, relative to people from larger hometowns (Balas & Saville, 2015). A follow-up study replicated the hometown effect in the CFMT, but it was not generalized to a card-sorting task (Balas & Saville, 2017). Participants from small hometowns, while poorer at grouping images of the same face together (and therefore creating more groups), actually made fewer misidentification errors within each group of identities (Balas & Saville, 2017). These mixed results suggest a more complicated mechanism to explain differences in face-recognition ability, with the possibility that it may depend on the task.

Indeed, a replication of the effect of hometown density on CFMT performance (Sunday, Dodd, Tomarken, & Gauthier, 2018) also found the effect did not replicate to another face task. In that study we aimed to explore whether the effect of hometown population density on CFMT performance would extend to other tasks that share one of two properties with the CFMT, namely (i) whether it used faces, and (ii) whether it used a learning format where a few items are tested repeatedly. To address these aims, participants were tested with the Vanderbilt Face Matching Test (VFMT, Sunday, Lee, & Gauthier, 2018), a face test like the CFMT but where each trial uses entirely new face identities (unlike the CFMT, in which six identities are studied and their recognition is tested on a series of trials). They were also tested on tasks with non-face objects (birds, mushrooms, cars and planes), that share the overall format with the CFMT in asking participants to study six identities from a given category and then testing recognition of those identities over several trials. When comparing 23 people from a small hometown to 84 people from a larger hometown, the only test to show a hometown effect was the CFMT. These results offered evidence that the advantage of coming from a large hometown applies to faces and not non-face objects, but it also suggested that not all face recognition tasks would be equally sensitive to this effect.

In the present work, we hypothesized that because the CFMT requires learning across trials whereas the VFMT does not, the process sensitive to early experience may be relatively specific to the requirements of a test that relies more on long-term face memory. Face memory and face perception have been dissociated to some extent in prior work (De Heering et al., 2014; Dalrymple, Garrido, & Duchaine, 2014; Weigelt et al., 2014). Face memory matures significantly later – after age ten – than face perception, which matures before the age five (Weigelt et al., 2014). Dalrymple et al. (2014) found that performance on the CFMT and the Cambridge Face Perception Test (Duchaine, Germine, & Nakayama, 2007) dissociate in adults with developmental prosopagnosia. In adult bilateral congenital cataract patients, performance on tasks requiring the use of face memory was not correlated with performance on perceptual tasks (De Heering et al., 2014). Indeed, the presence of a memory component in a paradigm has been shown to influence how tests relate to one another. Specifically, Richler, Floyd, and Gauthier (2015) demonstrated that when there is a learning component common to both a holistic processing measure and the CFMT, the tasks correlate whereas when the learning component is removed from the holistic processing measure, the correlation becomes non-significant.

Here we explored further whether the hometown effect is only present in tests of face recognition ability that contain a learning component, by using tests in multiple formats. Specifically, we used six face recognition measures: three with learning components (the CFMT and two similarly-formatted measures) and three without learning components (the VFMT, a face-part processing measure, and a face ensemble perception measure). If we find evidence for this hypothesis, it would suggest that the hometown effect is related to face memory rather than face perception and recognition. More generally, by using several face recognition measures, we aimed to better characterize how well the hometown effect generalizes across different tasks.

A second goal of this work was to obtain a more balanced sample of individuals coming from very small vs. larger hometowns. Using such a sample, we hoped to have more sensitivity to look at performance over the continuous range of log population density (hereafter logPD), as well as investigate how gender interacts with population density to influence performance on face tasks. These are questions that were difficult to address in our prior work, as well as in the studies by Balas et al., which had smaller samples (Balas & Saville, 2015: N = 37 total, Balas & Saville, 2017: N = 39). Aside from the overall sample size, these two studies had only 19 and 21 participants respectively in their small hometown group, and while (Sunday, Dodd et al., 2018; Sunday, Lee et al., 2018) had a total sample of 107, still only 23 participants fell into the "small hometown" group. In the (Sunday, Dodd et al., 2018; Sunday, Lee et al., 2018) study, the sampling strategy was simply to recruit individuals from the University of Nebraska-Lincoln campus, expected to include many individuals from small hometowns. However, the observed distribution of log population density was heavily right skewed, with the majority of the participants still coming from a large hometown. For this reason, using population density as a continuous variable was not useful and we had to resort, as in previous work, to less than ideal dichotomization of the sample into low and high population density groups. For the present work, we decided to explicitly recruit more people from small hometowns using special flyers and succeeded in obtaining a much more even distribution for log population density (Fig. 1), with 50 of the 90 participants falling into what would have been the "small hometown" group using criteria from prior work. More importantly, sampling more evenly across the range of log population density allowed more sensitive continuous analyses (MacCallum, Zhang, Preacher, & Rucker, 2002), whereby we did not have to set an arbitrary criterion to separate small vs. large hometowns.

In addition to looking at gender as a moderating variable, we also explored possible mediators of hometown effects. We assessed extroversion using questions from the IPIP (Goldberg et al., 2006) and trait anxiety by using questions from the STAI (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), to explore factors other than experience *per se* that could vary as a function of hometown size and mediate the effects previously reported. Some work has identified a relation between face recognition and self-reported extraversion (Lander & Poyarekar, 2015; Li et al., 2010) and between face recognition and anxiety (Megreya & Bindemann, 2013; Mueller, Bailis, & Goldstein,

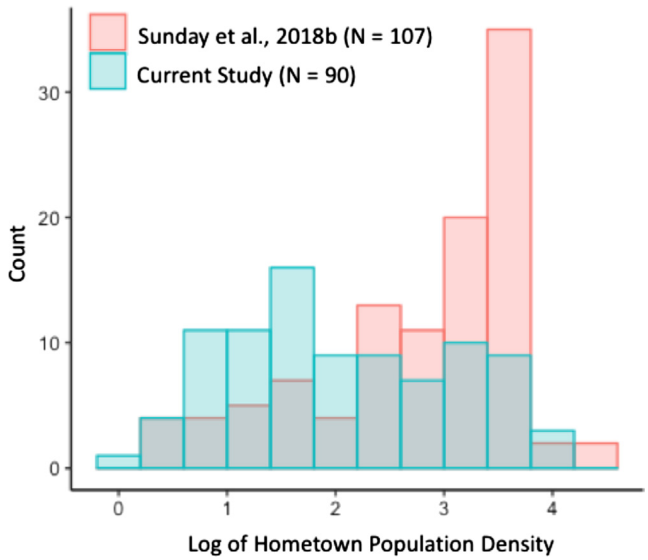

Fig. 1. Histograms of the log of hometown population densities for Sunday, Dodd et al. (2018) (N = 107) and the current study (N = 90).

1979). Finally, we sought to obtain more detailed information regarding the population densities of participants' previous residences and the ages during which they resided there, to see if we could improve on our characterization of our participants' experience with faces.

To foreshadow our main results, unlike prior work, we found no evidence of a main effect of population density on the CFMT or on any other face recognition task. However, we find evidence that gender interacts with population density to predict performance across several face recognition tasks. Extroversion and trait anxiety showed no clear relation with hometown population density or face recognition.

2. Methods

2.1. Participants

Our final sample included 90 participants, recruited from the University of Nebraska-Lincoln campus via posted flyers. A total of 148 participants responded to this flyer over the span of eight months. Because research suggests face-recognition ability peaks around age 30 (Germine, Duchaine, & Nakayama, 2011), here we excluded three individuals over the age of 30. To minimize variance associated with differences in face recognition due to the other-race effect (De Heering et al., 2010; Sangrigoli et al., 2005; Yovel et al., 2012), and because it was difficult to find hometown population data for international residences, we excluded 23 participants who reported hometowns outside the United States. A further 32 participants were excluded because they did not complete the hometown questionnaire. Thus, we analyzed data from the remaining 90 participants (29 males, 61 females; Age: Mean = 20.84, SD = 2.82, Range = 18–30; Ethnicities: Caucasian = 79, Asian = 5, Hispanic/Latino = 4, Pacific Islander = 1, Other = 1). However, of these participants, two did not complete the VET-Male, two did not complete the VHPT, three did not complete the VET-Female, and one did not complete the EP task. Upon completion of six tasks and a questionnaire, participants were compensated $22.50. Informed consent was obtained, and all research was conducted in compliance with both Vanderbilt and UNL Institutional Review Boards and the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Procedure

Upon contacting us, participants were allowed one week to individually complete six online tasks at any time in the following order: CFMT, VFMT, Vanderbilt Expertise Test – Caucasian males (VET-Male), Vanderbilt Holistic Processing Test for faces (VHPT-F), Vanderbilt Expertise Test – Caucasian females (VET-Female), and an ensemble perception face task (EP-Face), see Fig. 2. Three of these tasks are learning tasks that use a similar format with six studied targets: the CFMT, VET-Male and VET-Female. The other three tasks do not include a learning component: two of them use different face stimuli on every trial, the VFMT and the VHPT-F, and a third one, the ensemble perceptual face task, uses the same small set of morphed stimuli for the entire set of trials.

Trait anxiety questions from the State-Trait Anxiety Inventory (STAI) and items measuring extroversion from the International Personality Item Pool (IPIP) were collected at the end of the VET-Male and VET-Female, respectively. These tasks were administered online and took a total of approximately 1.5 h to complete. Participants were also asked to complete a questionnaire that consisted of six questions intended to obtain information about participants' hometowns and the variety and number of people with whom they may have interacted, discussed in detail below.

2.3. CFMT

The CFMT+ (Russell et al., 2009) is a learning-based face recognition task that entails three progressively more difficult phases consisting of eighteen, thirty, and fifty-four trials, respectively. In the introductory learning phase, participants study six Caucasian male exemplar faces presented individually in three alternative view-points for three seconds each. In the next three trials participants select the target face presented in the same viewpoint as the face that was studied from two distractor faces. Following a review period in which participants study the exemplar faces for twenty seconds, thirty trials are presented in a manner identical to that of the previous phase except with added variation in lighting and viewpoint to the faces. The third phase is identical except for the addition of Gaussian noise to the response trials. To index face recognition from the CFMT, percent accuracy was calculated (chance = 33%). The test took approximately 15 min to complete.

2.4. VFMT

The VFMT was designed to measure face recognition without requiring learning across trials (Sunday, Lee, et al., 2018). On each trial, participants study two unfamiliar Caucasian faces and then determine which one of the three proceeding faces matches one of the two studied faces. There are three practice trials followed by ninety-seven trials. Trials showing all male or all female faces are interleaved, with the exception of two catch trials, in which the two distractor faces differ in gender from the two studied faces. Responses were unspeeded and participants received no feedback. To index face recognition from the VFMT, percent accuracy was calculated (chance = 33%). The test took approximately 15 min to complete.

2.5. VET-Male and VET-Female

The VET-Male and the VET-Female are learning-based face recognition tasks designed to be similar in format to the CFMT (Ryan & Gauthier, 2016). The VET-Male uses only Caucasian male faces as stimuli and the VET-Female only Caucasian female faces. Both tasks are divided into three phases that entail six, six, and thirty-six trials, respectively. In the first phase, participants study six different Caucasian male or female exemplar faces for twenty seconds and then complete six trials requiring identification of the image that is identical to one of the studied faces. For the second phase, participants once again study the same six exemplar faces for twenty seconds. Participants then complete six similar response trials. The third phase consists of participants

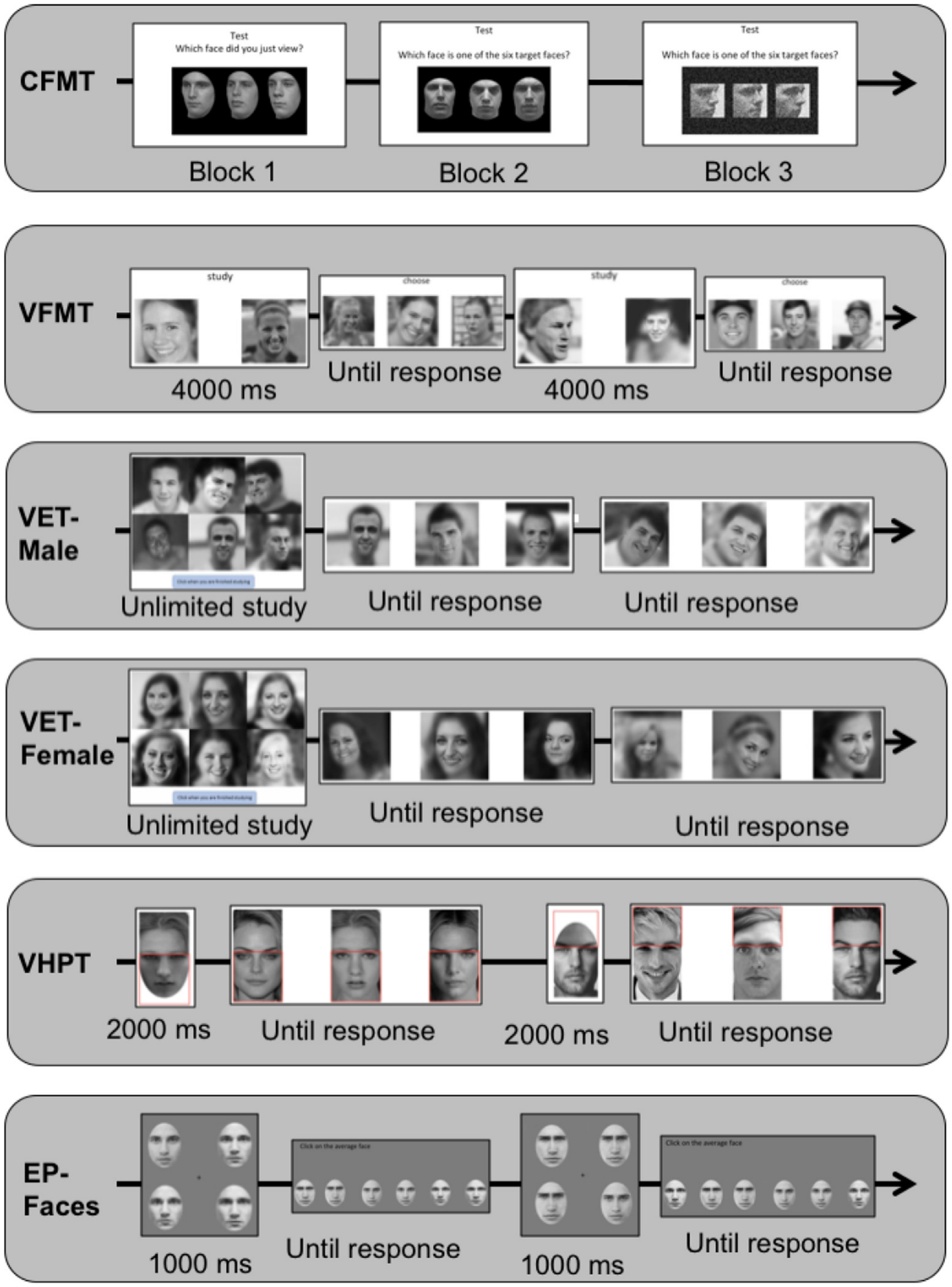

Fig. 2. Examples of task trials.

studying the six exemplar faces and then completing thirty-six response trials (three catch trials are included with significantly older distractors), where the correct images are no longer identical to those studied. Responses were unspeeded and participants only received feedback in the first two phases in the form of onscreen text. To index face recognition from the VET-Male and VET-Female, percent accuracy was calculated (chance = 33%). Each test took approximately ten minutes to complete.

2.6. VHPT

The VHPT consists of 180 trials designed to test individual ability to holistically process faces (Richler et al., 2015). Holistic processing, here operationalized as the failure to selectively attend to face parts, is considered a hallmark of expertise (Richler, Cheung, & Gauthier, 2011). In the present study we report both the index of holistic processing and the face part matching scores from this test, referred to as VHPT-HP and VHPT-Total respectively. In the task, participants were provided two seconds to study a particular section of the face and were then presented with a response trial that contained three faces. They then had to select the face that contained the task-relevant section while ignoring distractor sections. Half of the trials showed incongruent faces, which are those where the distractor sections are different from those presented during study. The other half showed congruent faces, where distractor sections are identical to those presented at study. Responses were unspeeded and participants received no feedback. Average percent accuracy of responses was determined separately for congruent and incongruent trials. To index holistic processing, the difference between congruent and incongruent percent accuracies were calculated, with larger differences indicating more holistic processing. To index face recognition from the VHPT, percent accuracy on all trials was calculated (chance = 33%). The test took approximately 25 min to complete.

2.7. EP-Face

The Ensemble-face (EP-Face) task measures a participant's ability to recognize the average of a set of faces with seventy-four trials. Using MorphAge software (version 4.2.4, Creaceed), six stimuli were generated by morphing three faces linearly to varying degrees (25% and 75%), producing six total face morphs. For each trial, following a one second presentation of a balanced Latin square arrangement of four faces, participants had to select the average of the four faces from a response display of six face morphs, which were presented in a linear sequence but with a different starting point on each trial. Responses were unspeeded and participants received no feedback. Each response was calculated by taking the absolute value of the difference between the response and the correct answer. These scores were then averaged to index performance, with higher scores indicating poorer performance (chance = 1.5). The test took approximately 15 min to complete.

2.8. STAI

The STAI is a self-evaluation questionnaire designed to measure both state and trait anxiety (Spielberger et al., 1983). For this study, only the twenty statements evaluating trait anxiety were presented following the VET-Male, since trait anxiety is more representative of general anxiety behavior that may occur during the face learning period. Participants rated on a scale from 1 (not at all) to 4 (very much so), whether the statements describe how they "generally feel". Since half of the statements reflect a lack of anxiety (e.g., "I feel secure"), these were reverse scored. Thus, higher total scores indicate higher levels of anxiety.

2.9. IPIP

The IPIP was created to be an open-source collection of questions that assessed personality based on the Big-Five factors: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Goldberg et al., 2006). Participants rated on a scale from 1 (very inaccurate) to 5 (very accurate), whether the statements describe themselves "as they generally are now." Because previous literature suggests a significant positive relationship between extraversion and face recognition ability, a relationship that is not mirrored for the other Big-Five factors (Li et al., 2010), only extraversion was measured for the purposes of this study. The relevant test items measuring for extroversion were included at the end of the VET-Female. Since half of the statements reflect introversion (e.g., "Keep in the background"), these were reverse scored. Therefore, higher total scores indicate greater extroversion.

2.10. Hometown questionnaire

Participants were asked to complete an open-ended questionnaire upon completion of six tasks. In the questionnaire, participants were asked, "Starting with your place of birth, please list the places you have lived, the zip code of this residence and the approximate ages you lived there." Additional information about the school size, frequency of vacations and family size was also collected but was not used due to discrepancies among participants in how the information was reported. Using the zip codes reported by the participants, population densities of participants' hometowns were determined from www.unitedstateszipcodes.org.

3. Results

Because we asked participants to list every place of residence rather than only their hometown (as in Balas & Saville, 2015; Sunday, Dodd et al., 2018), several of our participants reported multiple zip codes. We did this to have more information about participants' durations and ages of stay in each place of residence. However, preliminary analyses revealed that using only population density from the original hometown, the longest hometown, or a weighted averaged as a function of years spent in each location, yielded highly correlated population densities in this sample (r's = 0.58–0.89, $p = < .001$). Other work in China and Germany reported similar high correlations between hometown population size and indices of current urbanicity (Sindermann et al., 2017). Because it was unlikely that we could separate the contributions of early vs. later effects here, we simply chose to use the first reported place of residence as hometown, to be consistent with prior work. These analyses at least reveal that one should be cautious in judging the present effects to be effects of early experience. The population densities of each participants' first places of residence were positively skewed. Thus, we log transformed the data, similar to previous work (Sunday, Dodd et al., 2018). Data are available at https://figshare.com/articles/Gender_and_hometown_population_density_interact_to_predict_face_recognition_ability/7988138.

To index reliability for each test, Cronbach's α, a measure of internal consistency, was calculated for all tasks. These reliabilities, along with other descriptive statistics, are reported in Table 1. All tests showed acceptable internal consistencies, with the lowest reliability in the VHPT-HP (0.52).

All measures significantly positively correlate with one another, except for the VHPT-HP, consistent with prior work showing that this measure of holistic processing does not correlate with general face recognition (Konar, Bennett, & Sekuler, 2010; Richler et al., 2014, 2015). Note that the EP-Face task scores were reverse coded so that higher scores indicated better performance as in the other tests, avoiding negative correlations for simplicity. The Pearson correlations across tasks are reported in Table 2.

Surprisingly, logPD did not correlate with performance on any test (r's = −0.15–0.15, p's = .16–.95, Table 3). Additionally, no test correlated with either STAI or IPIP measures (STAI: r's = −0.17–0.05, p's = .13–.81; IPIP: r's = −0.14–0.16, p's = .15–.78, Table 3) and neither STAI nor IPIP correlated with logPD (STAI: $r_{86} = -0.08$, $p = .44$; IPIP: $r_{85} = 0.07$, $p = .53$).

To make sure that the failure to find a hometown effect, especially where it has been previously reported on the CFMT, is not due to using logPD as a continuous measure, we also verified that none of the tasks showed a hometown effect when the sample was split into small and large hometown groups by the same criterion as in Sunday, Dodd et al. (2018). Using a criterion of ≤85 ppl/mi^2 for small hometowns and > 85 ppl/mi^2 for large hometowns, none of the tasks showed a difference between groups, even in one-tailed tests (CFMT: $t(88) = -1.26$, $p = .11$; VFMT: $t(88) = -0.62$, $p = .27$; VET-Male: $t(86) = -0.22$, $p = .42$; VHPT-Total: $t(86) = 0.27$, $p = .40$; VHPT-HP: $t(88) = 0.57$, $p = .28$; VET-Female: $t(85) = 0.72$, $p = .24$; EP-Face: $t(87) = 0.57$, $p = .29$).

Table 1
Descriptive statistics for each behavioral test.

Task	N	Mean (SD)	Range	Reliability (α)
CFMT	90	0.55 (0.12)	0.62	0.85
VFMT	90	0.59 (0.09)	0.47	0.77
VET-Male	88	0.81 (0.12)	0.67	0.88
VHPT-Total	88	0.64 (0.08)	0.51	0.89
VHPT-HP	88	13.38 (8.18)	42.22	0.52
VET-Female	87	0.81 (0.16)	0.75	0.92
EP-Face	89	1.16 (0.24)	1.29	0.79

Table 2
Correlations of behavioral tests with 95% confidence intervals in parentheses. For r's > 0.31, *p*'s < .001, for r's > 0.28, *p*'s < .01. Correlations disattenuated for measurement error are reported in the upper-right in shaded region.

Task	CFMT	VFMT	VET-Male	VHPT-Total	VHPT-HP	VET-Female	EP-Face
CFMT	-	0.56	0.54	0.33	-0.03	0.61	0.48
VFMT	0.45 (0.27, 0.60)	-	0.58	0.46	0.00	0.66	0.52
VET-Male	0.46 (0.28, 0.61)	0.50 (0.33, 0.65)	-	0.54	0.06	0.79	0.35
VHPT-Total	0.29 (0.08, 0.47)	0.40 (0.21, 0.56)	0.47 (0.29, 0.62)	-	0.29	0.62	0.52
VHPT-HP	-0.02 (-0.23,0.19)	-0.00 (-0.21, 0.21)	0.04 (-0.17, 0.25)	0.20 (-0.01, 0.39)	-	0.04	-0.31
VET-Female	0.54 (0.37, 0.68)	0.58 (0.43, 0.71)	0.70 (0.58, 0.79)	0.55 (0.38, 0.68)	0.03 (-0.18, 0.24)	-	0.38
EP-Face	0.39 (0.20, 0.55)	0.42 (0.24, 0.58)	0.29 (0.08, 0.47)	0.43 (0.24, 0.59)	-0.20 (-0.39, 0.01)	0.31 (0.10, 0.49)	-

Table 3
Correlations of behavioral tests with log of population density, gender, IPIP and STAI. 95% confidence intervals in parentheses. No correlations are significant except VET-Female with gender (*p* < .01), VFMT with gender (*p* < .01), and VHPT-Total with gender (*p* < .05).

Task	Log PD	Gender	IPIP	STAI
CFMT	−0.15 (−0.35, 0.06)	−0.21 (−0.40, 0.00)	0.11 (−0.10, 0.31)	−0.14 (−0.34, 0.07)
VFMT	−0.03 (−0.24, 0.18)	−0.30 (−0.48, −0.10)	0.16 (−0.06, 0.35)	0.05 (−0.16, 0.26)
VET-Male	−0.01 (−0.22, 0.20)	−0.19 (−0.38, 0.02)	0.03 (−0.18, 0.24)	−0.16 (−0.35, 0.06)
VHPT-Total	0.06 (−0.16, 0.26)	−0.27 (−0.46, −0.07)	0.12 (−0.09, 0.33)	−0.07 (−0.27, 0.15)
VHPT-HP	−0.06 (−0.26, 0.16)	−0.11 (−0.31, 0.10)	−0.14 (−0.34, 0.07)	0.04 (−0.17, 0.25)
VET-Female	0.15 (−0.06, 0.35)	−0.28 (−0.46, −0.08)	0.05 (−0.16, 0.26)	−0.17 (−0.36, 0.05)
EP-Face	−0.02 (−0.22, 0.19)	−0.12 (−0.32, 0.09)	0.10 (−0.11, 0.31)	0.03 (−0.18, 0.23)

Next, we asked whether gender interacted with population density to predict performance on each of the tasks. We entered gender (dummy coded), logPD and their interaction into multiple regressions, predicting each task separately (Fig. 3, Table 4).

All tasks designed to measure face recognition show a similar interaction (Fig. 3) even though the interaction was significant for three tasks (CFMT, VFMT, VET-Female), approaching the 0.05 level for two tasks (VHPT-Total and EP-Face), and far from significant in the last task (VET-Male). Unlike the other six indices, the VHPT-HP measures holistic processing and thus we would not expect it to show the same pattern as the six face recognition indices. For that task, the multiple regression model showed no effect of logPD ($t(87) = 0.001$, $p = .99$, $r = -0.06$), a small and non-significant advantage for women over men ($t(87) = 1.66$, $p = .10$, $r = 0.11$), and a non-significant interaction between gender and logPD ($t(87) = -1.34$, $p = .18$, $r = 0.02$, Adjusted $R^2 = 0.2\%$ of VHPT-HP variance).

Because of robust correlations across six of the tasks (Table 2) and because they theoretically share common variance pointing to face recognition ability, we averaged together the z-transformed scores on each task to produce a less noisy estimate of face recognition. This follows the principle of aggregation (Rushton, Brainerd, & Pressley, 1983), whereby "the sum of a set of multiple measurements is a more stable and representative estimator than any single measurement". The multiple regression model accounted for 15.4% of the variance (Adjusted R^2) in overall face recognition, with no effect of logPD ($t(86) = 1.1$, $p = .27$, $r = 0.12$), an advantage for women over men ($t(86) = 4.08$, $p < .001$, $r = 0.41$) and, most interestingly, a significant interaction between gender and logPD ($t(86) = 2.97$, $p = .004$, $r = 0.31$). As can be appreciated from the scatterplot (see Fig. 4) in small hometowns, women outperform men on face recognition and this effect disappears with increasing population density. The change in performance as a function of logPD is numerically larger (and significant) for men ($r = 0.42$, $p = .03$) than it is for women ($r = -0.22$, $p = .10$), thereby suggesting that men are driving the interaction (although a much larger sample study would be necessary to test if one effect is larger than the other).

4. Discussion

In this study we explored how hometown density relates to face recognition ability in several face recognition measures. Based on earlier findings of a hometown effect for the CFMT and not the VFMT, we predicted that face recognition measures that involved learning (CFMT, VET-Male and VET-Female) would vary as a function of hometown density whereas the three face recognition measures that did not involve learning (VFMT, EP-Face and VHPT-Total) would not. We found little support for this prediction and in fact, performance on none of our tasks showed a main effect of hometown density. This is despite having more participants from relatively small hometowns (≤85 ppl/mi^2) than in our previous study (current study N = 50, previous work N = 23). This allows us to reject the hypothesis, formulated on the basis of only two tasks showing diverging results in (Sunday, Dodd et al., 2018; Sunday, Lee et al., 2018)), that the hometown effect may be specific to face memory (or to be more precise, tasks that rely on long-term memory).

Instead, we observed an interaction between gender and hometown density, with a pattern consistent (although not significant in all cases) across six face recognition measures. Females from small hometown outperformed males from smaller hometowns but no such difference was observed in subjects from larger hometowns. As best we can tell from the present results, the interaction is driven by higher face recognition performance for men as a function of PD.

The present results highlight the importance of using a range of tasks in the study of individual differences. When only two tasks are

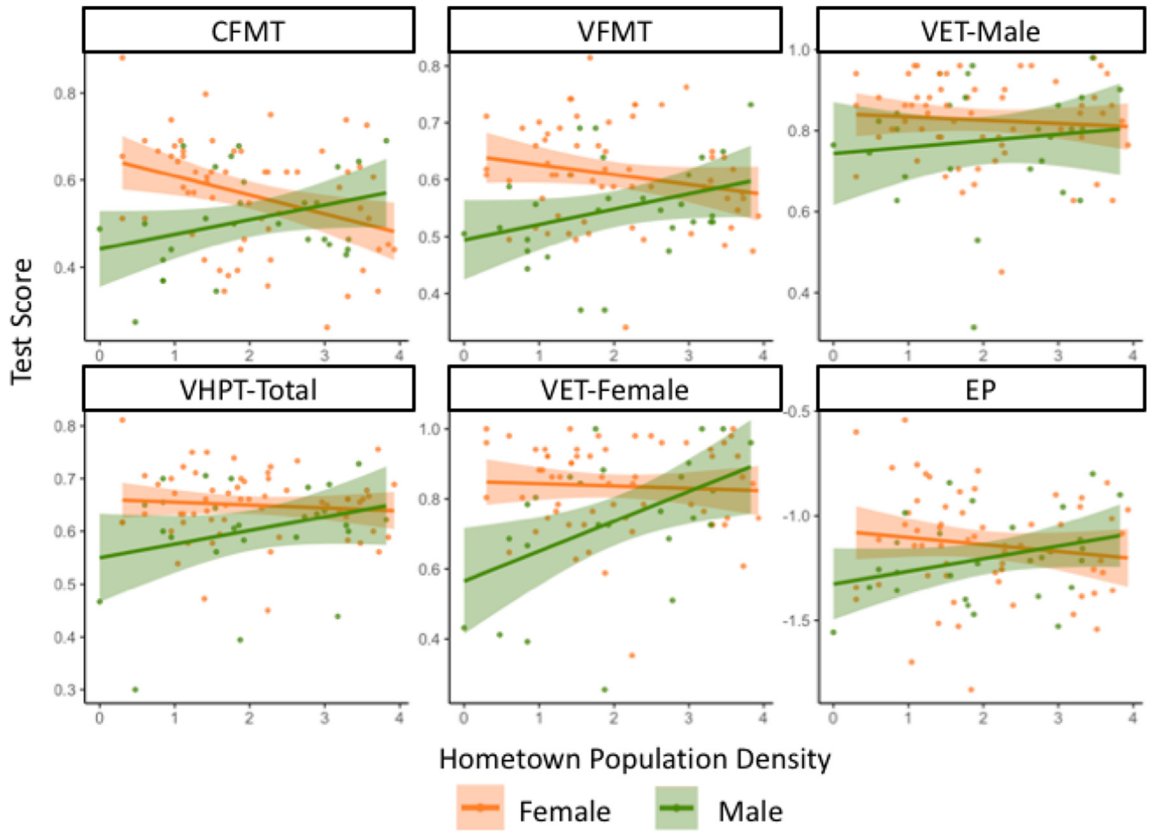

Fig. 3. Scatterplots of scores for each test indexing face recognition ability by hometown density (logPD). Females (orange) and males (green) are shown with separate regression lines and shaded regions indicated 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Results of the multiple regressions analyses.

Test	Predictor variable	*Estimate (SE)*	*T*	*p*
CFMT	LogPD	−0.01 (0.01)	−0.39	.70
	Gender	0.11 (0.03)	−3.64	< .001
	LogPD*Gender	−0.04 (0.01)	3.03	.003
	Adjusted R^2: 0.13			
VFMT	LogPD	0.01 (0.01)	0.51	.61
	Gender	0.08 (0.02)	−3.47	< .001
	LogPD*Gender	−0.02 (0.01)	2.35	.02
	Adjusted R^2: 0.12			
VET-Male	LogPD	0.004 (0.01)	0.29	.77
	Gender	0.05 (0.03)	−1.62	.11
	LogPD*Gender	−0.01 (0.01)	0.90	.37
	Adjusted R^2: 0.01			
VHPT-Total	LogPD	0.01 (0.01)	1.21	.23
	Gender	0.06 (0.02)	−2.88	.005
	LogPD*Gender	−0.02 (0.01)	1.87	.07
	Adjusted R^2: 0.08			
VET-Female	LogPD	0.04 (0.02)	2.47	.02
	Gender	0.14 (0.04)	−3.89	< .001
	LogPD*Gender	−0.05 (0.02)	2.89	.005
	Adjusted R^2: 0.16			
EP-Face	LogPD	−0.01 (0.03)	−0.52	.61
	Gender	−0.13 (0.06)	−2.11	.03
	LogPD*Gender	0.05 (0.03)	1.79	.08
	Adjusted R^2: 0.02			

contrasted, it is possible to make too much of a difference in pattern between them (a similar argument has been made previously regarding over-interpreting the difference between one face task and one object task, see Gauthier, 2017). Here, by using six different tasks, we can see a pattern that was fairly general across face tasks, whether they rely on long-term memory or not, whether they use many stimuli or not, whether judgments are about parts, wholes or averages. While there were differences across tasks, for instance the VET-Male did not show a significant interaction between hometown density and gender, the overall pattern of results suggests we should be careful not to interpret this as a difference driven by task specifics. Indeed, it is not the only face learning task, nor the only task that uses male faces. Given all this, the most careful conclusion may be that an interaction between gender and hometown density is observed for a task-general face recognition factor that contributes to many face tasks. Given the similar pattern of interaction observed across tasks, it would require a very large sample to test the hypothesis that this interaction is larger for one task than another.

We should, however, be careful to assume that this interaction will generalize to all other samples. First, we made a special effort to recruit more individuals from small hometowns, in addition to recruiting in an area where such individuals are more numerous than in many other places in the U.S. In addition, by using students from the University of Nebraska-Lincoln, we may have inadvertently selected for specific characteristics that modulate one's face recognition ability differently depending on gender. Female participants from small hometowns wishing to pursue higher education may have some quality not present in their male counterparts, and this small hometown female sample may not be representative of females from any smaller U.S. hometowns.

There are numerous differences associated with growing up in rural vs. urban environments. For instance, those in urban areas maybe have more access to medical treatment and to more diverse cultural activities, whereas those in more rural areas may be less affected by traffic, pollution and spend more of their time in natural settings. The latter in particular may benefit executive attention (Gamble, Howard, &

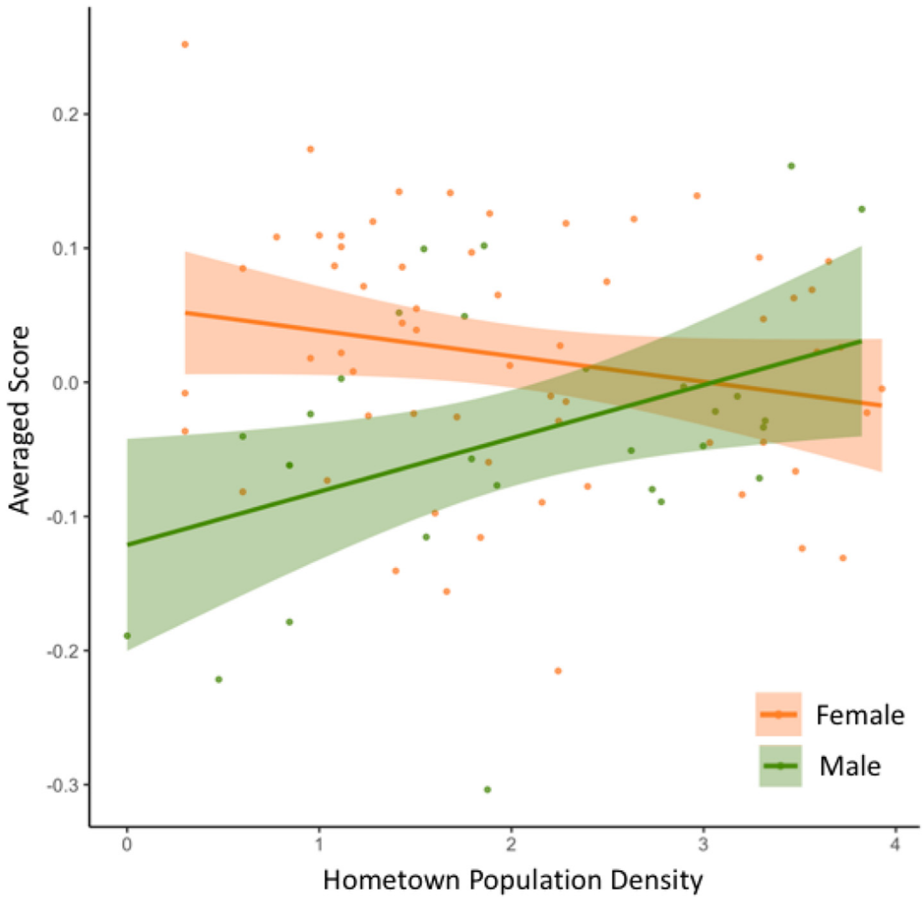

Fig. 4. Scatterplots of average z-scored performance on all 6 measures by hometown density (logPD). Females (orange) and males (green) are shown with separate regression lines and shaded regions indicated 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Howard, 2014; Schutte, Torquati, & Beattie, 2017). Moreover, where people live is a predictor of mental health, with depression and schizophrenia more prevalent in individuals who live in urban than rural conditions (Pedersen & Mortensen, 2001). Others have theorized that growing up in these different environments, and interaction between rural/urban differences with genetic effects, could have a lasting impact on important aspects of personality (Sindermann et al., 2017). Such effects could differ as a function of gender, for instance these authors suggested that in a Chinese sample, living in larger cities might buffer women against negative emotions through more access to social and employment opportunities. In the same sample, a correlation between PLAY (a primary emotional system involved in learning social competencies, see Montag & Panksepp, 2017) and urbanicity was observed especially among males. This suggests that perhaps the improvement in males from larger hometowns observed here could reflect differences in social competencies between rural and urban regions, though this would be purely speculative given it is unclear if such a finding could even be replicated in a Western country like the United States. Gender × urbanicity interactions on emotional traits found in a Chinese sample were not observed in a German sample (Sindermann et al., 2017). This illustrates the challenges of studying the interaction of gender with population density, a factor that may be associated with very different physical and social environments in different parts of the world and even within the U.S. One of several possibilities is that males who live in small towns may have fewer opportunities to interact with others whereas males in larger cities may be unable to avoid seeing more people. In that sense, hometown PD could be a proxy for the quality and number of a person's opportunity for social interaction. Prior work found that the quality of interactions with other-race individuals mitigates the own-race bias (Walker & Hewstone, 2006).

One perhaps easier question to answer is whether the interaction we observe generalizes from this specific sample to other samples from the same population. To address this, we re-analyzed the results from the two face recognition tasks in Sunday, Dodd et al. (2018). Because that sample was highly skewed towards higher population density, no hometown effect was observed when using population density as a continuous measure, although they were present in the dichotomized groups, and gender effects were not investigated. In light of the current results, we asked whether the same pattern we observed here was present in these data. Details of that study can be found in Sunday, Dodd et al. (2018). Briefly, data for 107 subjects were available for analysis – here we use logPD for hometown, as measured in the current study, and data from the CFMT+ and VFMT. Multiple regression was used to look for effects of gender, logPD and their interaction. Hometown population density was the only significant predictor of CFMT performance. A gender effect was however observed for the VFMT, as was an interaction between gender and hometown population density (Table 5, Fig. 5).

Table 5
Results of the multiple regressions analyses.

Test	Predictor variable	*Estimate (SE)*	*T*	*p*
CFMT	LogPD	0.02 (0.01)	1.67	.10
				.70
	Gender	0.02 (0.04)	0.537	.59
	LogPD*Gender	0.003 (0.01)	0.25	.80
	Adjusted R^2: 0.05			
VFMT	LogPD	0.02 (0.01)	1.61	.11
				.61
	Gender	0.02 (0.04)	2.43	.02
	LogPD*Gender	0.003 (0.01)	−2.04	.04
	Adjusted R^2: 0.04			

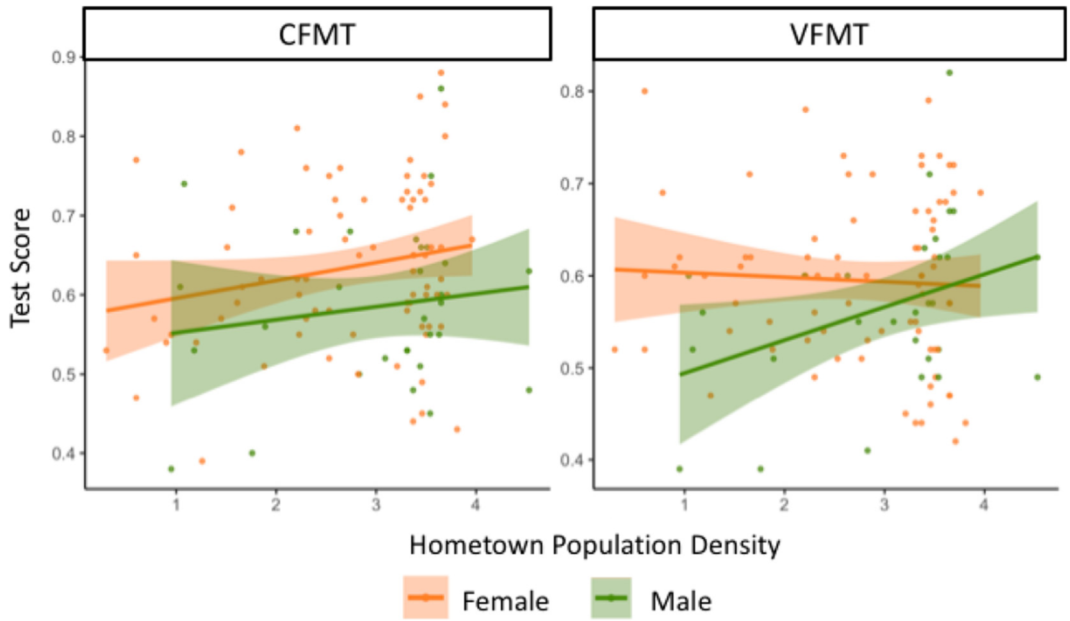

Fig. 5. Scatterplots of scores for each test by hometown density *(log PD). Females (orange) and males green) are shown with separate regression lines and shaded regions indicated 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. General discussion

This work used oversampling methods to study the relation between hometown population density and face recognition ability. We first tested, and found no support for, the hypothesis that hometown population density related to face recognition memory and not perception. However, across most of our face recognition tasks, and in the common variance across them, we found an interaction between gender and hometown population density on all tasks. Specifically, we observed that in individuals from less dense hometowns, females perform better than males, but this advantage was diminished or disappeared in more urban hometowns. A previous dataset from Sunday, Dodd et al. (2018) revealed a similar interaction in the VFMT, but not the CFMT. Across both studies, the interaction with gender was obtained only when a gender effect was also obtained.

A gender effect is not always observed on face tasks (Duchaine & Nakayama, 2006; Ryan & Gauthier, 2016) although when it is obtained it has consistently been an advantage for women (Goldstein & Chance, 1970; Lewin & Herlitz, 2002; Lovén et al., 2013; Rehnman & Herlitz, 2007). To be clear, there does not appear to be a simple way to predict when to expect this female advantage in face recognition, despite intuitions that aspects of the task or the specific sample may be important. For instance, the same task, like the CFMT, can show this effect in one study (the present results) and not another (Sunday, Dodd et al., 2018), despite the two studies recruiting from the same population (see Stanley & Spence, 2014, for a reminder that measurement error alone is more than sufficient to explain differences across replications). In the present study, the same participants showed a female advantage on some tasks (the CFMT and VET-Female) but not a strikingly similar task (the VET-Male). The only study to report a male advantage for a category of faces (Transformer faces) linked gender effects in face recognition to differences in experience with different kinds of faces (Ryan & Gauthier, 2016) – but here, the VET-Male and CFMT both used Caucasian male faces and showed different results.

Our results speak to the importance of including multiple measures for the interpretation of observed results. Less than 20 years ago, we did not know that people vary as much as they do in face recognition ability. This knowledge came about through the creation of tasks, like the CFMT, that are sensitive to the range of performance in the normal population (Duchaine & Nakayama, 2006). Since then, many other tasks sensitive to a similar range of ability and with good psychometric properties have been developed, leading to a range of new findings, including that face recognition varies with age (Germine et al., 2011), has a strong genetic component (Shakeshaft & Plomin, 2015; Wilmer et al., 2010), is influenced by experience (Balas & Saville, 2015, 2017; Devue & Barsics, 2016; Ryan & Gauthier, 2016; Wan, Crookes, Reynolds, Irons, & McKone, 2015) and is not related to holistic processing (Richler et al., 2015). The development of such tasks, suitable for individual differences measurement, by employing distinct stimuli, using different formats, and exploiting different aspects of face recognition (as well as non-face object recognition, Richler et al., 2019), is critical for a comprehensive interpretation of this body of work. Using latent variable modeling in future studies, which allows researchers to estimate latent variables such as underlying abilities, through the common variance from several indicators (for instance, many different tasks), may be particularly productive (Tomarken & Waller, 2005, see Richler et al., 2019). As of now, our results point to a general pattern of interaction between hometown size and gender, on the variability in a face processing ability that contributes to performance regardless of the specific task. We did not observe a sufficiently consistent pattern of differences across task to narrow it down to a more specific ability.

Ultimately, questions remain regarding the nature of the hometown effect. Future studies should consider evaluating if there are selection biases that may lead to differences in samples for small and large hometowns, such that one gender sample is inherently advantaged in face recognition. Other factors, particularly cultural differences between small and large hometowns, also represent interesting avenues of investigation. When it comes to variability in face recognition in the normal population, many important factors may be currently unsuspected. For instance, recent work suggests that bilingualism influences relative performance on versions of the CFMT for own vs. other race faces (Burns, Tree, Chan, & Xu, 2019). While the bilingualism effect is interesting, it will likely require several studies, in several populations and with additional tests, to understand the nature of the result. In the case of the hometown effect, which was originally proposed to index childhood experience with faces, our work suggests that special sampling efforts will be needed to collect data suitable to disentangle whether the hometown effect is truly an effect of early experience with faces, a more general effect of living in urban vs. rural areas, and what aspects of this experience influence face recognition.

The original hometown study (Balas & Saville, 2015) also found an effect on the N170 potential, and it will be interesting to ask whether the gender × hometown interaction found here is also obtained on this neural measure.

Acknowledgements

This work was funded by the Temporal Dynamics of Learning Center (National Science Foundation Grant SMA-1640681 and BCS-153486). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (1445197). We would like to thank Jackie Floyd for her help collecting data for this study.

References

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia, 69*, 211–217.

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research, 141*, 228–236.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*(6), 248–254.

Burns, E. J., Tree, J., Chan, A. H. D., & Xu, H. (2019). Bilingualism shapes the other race effect. *Vision Research, 157*, 192–201.

Dalrymple, K. A., Garrido, L., & Duchaine, B. C. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10–20.

De Bruïne, G., Vredeveldt, A., & van Koppen, P. J. (2018). Cross-cultural differences in object recognition: Comparing asylum seekers from Sub-Saharan Africa and a matched Western European control group. *Applied Cognitive Psychology, 32*(4), 463–473.

De Heering, A., Bracovic, A., & Maurer, D. (2014). Starting school improves preschoolers' ability to discriminate child faces. *Ecological Psychology, 26*(1–2), 16–29.

De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science, 13*(1), 181–187.

Devue, C., & Barsics, C. (2016). Outlining face processing skills of portrait artists: Perceptual experience with faces predicts performance. *Vision Research, 127*, 92–103.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology, 24*, 419–430.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*, 576–585.

Gamble, K. R., Howard, J. H., & Howard, D. V. (2014). Not just scenery: Viewing nature pictures improves executive attention in older adults. *Experimental Aging Research, 40*(5), 513–530.

Gauthier, I. (2017). Domain-specific and domain-general individual differences in visual object recognition. *Current Directions in Psychological Science, 27*(2), 97–110.

Germine, L., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201–210.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Goldstein, A. G., & Chance, J. E. (1970). Visual recognition memory for complex configurations. *Perception & Psychophysics, 9*(2), 237–241.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., & Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences, 111*(2), 823–828.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*(1), 38–43.

Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 811–817.

Lander, K., & Poyarekar, S. (2015). Famous face recognition, face matching, and extraversion. *Quarterly Journal of Experimental Psychology, 68*(9), 1769–1776.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—Women's faces make the difference. *Brain and Cognition, 50*(1), 121–128.

Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology, 3*(4), 295–298.

Lovén, J., Svärd, J., Ebner, N. C., Herlitz, A., & Fischer, H. (2013). Face gender modulates women's brain activity during face encoding. *Social Cognitive and Affective Neuroscience,* nst073.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research, 69*, 10–22.

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology, 25*(1), 30–37.

Montag, C., & Panksepp, J. (2017). Primary emotional systems and personality: An evolutionary perspective. *Frontiers in Psychology, 8*, 464.

Mueller, J. H., Bailis, K. L., & Goldstein, A. G. (1979). Depth of processing and anxiety in facial recognition. *British Journal of Psychology, 70*(4), 511–515.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*, 1321–1323.

Pedersen, C. B., & Mortensen, P. B. (2001). Evidence of a dose-response relationship between urbanicity during upbringing and schizophrenia risk. *Archives of General Psychiatry, 58*, 1039–1046.

Picci, G., & Scherf, K. S. (2016). From caregivers to peers: Puberty shapes human face perception. *Psychological Science, 27*(11), 1461–1473.

Putzar, L., Hötting, K., & Röder, B. (2010). Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restorative Neurology and Neuroscience, 28*, 251–257.

Rehnman, J., & Herlitz, A. (2007). Women remember more faces then men do. *Acta Psychologica, 124*, 344–355.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science, 22*, 464–471.

Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The vanderbilt holistic face processing test: A short and reliable measure of holistic face processing. *Journal of Vision, 1014*(11).

Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision, 15*(9), 15–112.

Richler, J. J., Tomarken, A. J., Sunday, M. A., Vickery, T. J., Ryan, K. F., Floyd, J. R., ... Gauthier, I. (2019). Individual differences in object recognition. *Psychological Review, 126*(2).

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*, 18–38.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16*(2), 252–257.

Ryan, K. T., & Gauthier, I. (2016). Gender differences in recognition of toy faces suggest a contribution from experience. *Vision Research, 129*, 69–76.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & De Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*(6), 440–444.

Sawada, R., Sato, W., Kochiyama, T., Uono, S., Kubota, Y., Yoshimura, S., & Toichi, M. (2014). Sex differences in the rapid detection of emotional facial expressions. *PLoS One, 9*(4), e94747.

Schutte, A. R., Torquati, J. C., & Beattie, H. L. (2017). Impact of urban nature on executive functioning in early and middle childhood. *Environment and Behavior, 49*(1), 3–30.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), 12887–12892.

Sindermann, C., Kendrick, K. M., Becker, B., Li, M., Li, S., & Montag, C. (2017). Does growing up in urban compared to rural areas shape primary emotional traits? *Behavioral sciences (Basel, Switzerland), 7*(3), 60.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications are yours realistic? *Perspectives on Psychological Science, 9*(3), 305–318.

Sunday, M. A., Dodd, M., Tomarken, A., & Gauthier, I. (2018). How faces (and cars) may become special. *Vision Research, 157*, 202–212.

Sunday, M. A., Lee, W.-Y., & Gauthier, I. (2018). Age-related differential item functioning in tests of face and car recognition ability. *Journal of Vision, 18*, 2.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31–65.

Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20*(4), 461–475.

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social–motivational component and derives entirely from lifetime perceptual experience. *Cognition, 144*, 91–115.

Weigelt, S., Koldewyn, K., Dilks, D. D., Balas, B., McKone, E., & Kanwisher, N. (2014). Domain-specific development of face memory but not face perception. *Developmental Science, 17*, 47–58.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., ... Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America, 107*(11), 5238–5241.

Wolff, N., Kemter, K., Schweinberger, S. R., & Wiese, H. (2014). What drives social ingroup biases in face recognition memory? ERP evidence from the own-gender bias. *Social Cognitive and Affective Neuroscience, 9*(5), 580–590.

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology: Human Perception and Performance, 38*, 285–289.