



Patrolling the Borders: Experience with Multiple Out-Group Members is Associated with the Own-Ethnicity Effect

Siegfried Ludwig Sporer¹ · Jürgen Gehrke¹ · Kimberley A. McClure²

Published online: 22 June 2020

© The Author(s) 2020

Abstract

The own-ethnicity effect (OEE) is a recognition deficit for faces of other ethnic groups compared to one's own group. Thirty-two border patrol (i.e., police) officers at a major international airport expected to have high frequency contact with multiple other ethnic group faces were compared with 32 bank employees and 64 students. German participants viewed out-group faces of African Americans, Mexican Americans/Mexican Nationals, and Turkish faces and German (in-group) faces in a standard recognition and a delayed-matching task. Generally, performance was better as a function of self-reported daily job contact for both tasks. Patrol officers were better with Black faces, while students were better with German faces for the standard recognition task. Although an OEE was present for all groups, it was attenuated for border patrol officers. Job contact with multiple out-groups was associated with face processing. Errors in recognition and matching faces of other ethnic groups pose problems for security in an increasingly global society. Customs agencies and security organizations employing an ethnically diverse workforce along with perceptual training for recognizing and matching out-group members will be better prepared to meet the demands of international security.

Keywords Border security · Face recognition · Face matching · Cross-race effect · Police as witnesses

As international travelers may have noticed, security concerns have risen worldwide. At passport control, or when entering security-sensitive government or business buildings, patrol officers and security personnel have to match the appearance of the person in front of them to a passport photograph. Likewise, when cashing a check, or applying for a loan, a bank teller routinely asks for photo identification to compare the appearance of the person in front of him or her with the photograph on the ID card. The cognitive processes involved in these comparisons may be more difficult when the persons whose faces are to be matched come from different ethnic groups than the personnel checking an ID. There is an extensive literature that has shown that when people are asked to recognize persons from other ethnic groups, recognition is often worse than with faces of their own ethnic group. This effect has been termed cross-race effect, own-race effect, or

own-race bias. We prefer the term own-ethnicity effect (OEE), or more generally, in-group face processing advantage, as recent studies have demonstrated that the effect may be a function of the construction of social subgroups (e.g., Chiroro et al. 2008), rather than “race” which is an ill-defined term, anyway (cf. Malpass, 1993). In-group advantages have also been shown for faces of one's own age (see the meta-analyses by Martschuk and Sporer 2018; Rhodes and Anastasi 2012) and one's own gender (Herlitz and Lovén 2013). The purpose of the current study is to consider the amount of contact and the number of other ethnic groups a person might encounter in their occupation and whether these experiences are associated with the OEE.

Although the phenomenon of an OEE was described early in the twentieth century (Feingold 1914), it was first systematically examined by Malpass and Kravitz (1969) and has since been replicated in more than 100 studies (e.g., Slone et al. 2000; Wright et al. 2003). Meta-analyses of the OEE reveal a robust and reliable phenomenon (Bothwell et al. 1989; Meissner and Brigham 2001; Tredoux et al. 2020). Notably, the majority of studies have only involved African American and European American participants having to recognize African American and European American faces (see Horry and Wright 2009, for exceptions). While the OEE may play a significant role in

✉ Siegfried Ludwig Sporer
siegfried.l.sporer@psy.jlug.de

¹ Department of Psychology and Sports Science, Justus-Liebig University Giessen, Otto-Behaghel Str. 10F, 35394 Giessen, Germany

² Western Illinois University, Macomb, IL, USA

criminal procedures when the witness of a crime and the culprit stem from different ethnic origins (Garrett 2011; Wells et al. 1998), its causal role in such archival analyses cannot be established unequivocally (see Wright 2006).

Theoretical Accounts of the Own-Ethnicity Effect

Several theories have emerged, trying to explain the OEE, from the development of face schemata (Goldstein and Chance 1980), physiognomic differences (Goldstein and Chance 1976) to the differential contact hypothesis (Chance et al. 1975). The contact hypothesis assumes that with increased contact with people from a specific ethnic group the recognition performance for faces from this group increases. Research testing the contact hypothesis is mixed with evidence in support (e.g., Chiroro and Valentine 1995; Wright et al. 2003) and against it (Ng and Lindsay 1994). When contact has been measured via questionnaires, only 1 to 2% of the variance in recognition performance is accounted for across studies (Meissner and Brigham 2001). However, when relative proportions of specific ethnic groups in geographic areas or suburbs are considered as indicators of contact, prediction improves substantially (Tredoux et al. 2020).

Exposure to, and Motivation to Recognize, Other Ethnicity Faces

The current research addresses whether quantity of contact with other ethnic groups facilitates processing of other ethnic faces, thereby attenuating the OEE. From a social psychological perspective, Malpass (1990) has provided an insightful discussion of the social utility inherent in in-group/out-group contact. For example, a Black student with a White teacher who marks his or her exams is likely to be motivated to recognize the teacher consistently. In considering the social utility of other ethnicity recognition, Dunning et al. (1998) tested the recognition performance of White basketball fans in the USA. As most of the professional basketball players are Blacks, White basketball fans (i.e., “experts”) were expected to show a greater interest in the players and therefore better recognition performance with Black faces compared to White non-basketball fans (i.e., “novices”). The results supported the social utility explanation of other ethnicity recognition. White expert basketball fans recognized Black faces better than White novices. Still, this experiment does not draw a distinction between the quality and the quantity of contact between different ethnic groups. It might be argued that White basketball fans also have greater contact with Blacks because they also play basketball more often and watch games live or on television. Hence, it is necessary to separate the two concepts of quality and quantity of contact

with other ethnic groups, as in most studies these concepts are confounded (see also Wright et al. 2003).

Slone et al. (2000) analyzed quality and quantity of contact of White participants with Blacks and identified six factors describing different aspects of quantity and quality of contact—present contact, public and personal non-intimate settings, past contact, business settings, intimate settings, and past friends. Of these dimensions, only “present contact” defined as the amount of everyday contact with Blacks on campus, in dorms, etc. correlated with recognition performance (but see McKone et al. 2019, who postulate a “critical period” for acquiring recognition expertise for other-race faces approximately under age 12).

Occupational Contact and the Own-Ethnicity Effect

The present study tests the assumption that contact will affect the OEE and extends the literature by considering contact with faces of four different ethnicities, and the frequency of contacts with these ethnic groups, within an occupational context. We tested border patrol officers, a special police force at the International Airport in Frankfurt, Germany, bank clerks at a bank in a small university town who were working in customer service,¹ and college students. It was anticipated that border patrol officers would have the most contact with people of other ethnicities and demonstrate the weakest OEE compared to bank tellers or students. Bank clerks, while likely to have contact with many people, are less likely to have contact with multiple ethnic groups and therefore serve as a reasonable control group. Similarly, students were expected to be a viable control group as they have daily contact with many other people, but less contact with students from other ethnic groups, and their “occupation” as a student provides little or no motivation to remember other ethnicity faces.

Border patrol officers at an international airport are more likely to encounter foreign faces representing all kinds of ethnic groups because of the number of international flights. This is particularly true of the Frankfurt airport which is one of the busiest hubs in the world for connecting international flights. Unfortunately, no statistics are made publicly available on the national origins and ethnic composition of passengers; however, it seems reasonable to assume that the work of border patrol officers, by definition, fosters encounters with people from other ethnic groups. Bank clerks as well as students, alternatively, are more likely to encounter German (White) faces than faces of other ethnic groups because of the composition of the general population as well as of the subpopulation of students consisting of a

¹ We are indebted to a reviewer of an earlier version of this manuscript, Dan B. Wright, for suggesting an additional comparison group with high-frequency contact of other people (but not necessarily of out-groups).

majority of White Germans. The Turkish population in Germany represents the largest proportion of persons with a migration background which make up approximately 25% of all residents in Germany (Statistisches Bundesamt 2019). The likelihood of daily encounters with Turks would be expected to be comparable for border patrol officers, bank clerks, and college students. Exposure to Turks would also be more likely than to persons from any other countries or continents.

Exposure to Multiple, Other Ethnic Groups

It is noteworthy that the majority of studies on the OEE have used faces of African Americans (Blacks) and European Americans (Whites), and a few studies have also investigated the OEE with faces of other ethnic groups, primarily Hispanics (Mexican American and Chinese; e.g., Platz and Hosch 1988; McKone et al. 2019).² Also, Sporer (1999) has shown an in-group advantage of German students for German compared to Turkish faces while there was no difference in performance among Turkish students who had lived most of their life in Germany. Furthermore, Sporer et al. (2007) have shown a similarly asymmetric OEE with Turkish and Austrian children in a matching task with faces of Turks and of Germans. While both groups of children took longer to match Turkish than German faces, the difference was much more pronounced with the Austrian children. To our knowledge, the majority of studies on the OEE involved two ethnic groups. We intended to extend this database by including faces from three ethnic groups which differ regarding the likelihood of contact in daily encounters.

In particular, we used faces representing Blacks, Hispanics, Turks, and Germans, essentially establishing multiple other ethnicity sub-categories to mimic the variation in ethnicity that is likely to be experienced in an international setting like the Frankfurt airport. We expected differences in the magnitude of the OEE as a function of the likelihood of contact with these groups. Overall, recognition performance should be best with White German faces, followed by Turkish, Hispanic, and Black faces. Most importantly, we expect the OEE to be attenuated for border patrol officers compared to bank clerks and students for Black, Hispanic, and Turkish faces.

Quality of contact was expected to be comparable for border patrol officers and bank clerks. And although the tasks differ (i.e., passport check vs. serving customers) they should not differ, for example, in the interest for a certain group (as the basketball fans did in the study by Dunning et al. 1998), and more generally, the social utility of these interactions should be comparable (Malpass 1990). Both occupational

contexts require short interactions with other persons across a counter or desk. Pre-experiment interviews and observations with border patrol officers showed that the average time for a passport check is about 7 s for each passenger. Thus, the average time an officer is in contact with a passenger is only slightly longer than in classic recognition experiments. Bank clerks, likewise, serve the customer over a counter, usually over a short period of time, but of course serving different purposes. Because of the nature of both occupations, they are task-oriented and are not likely to have any personal interest in the individual they are interacting with (except, perhaps, bank employees who are evaluating the credit worthiness of a customer).³ Regarding students, most encounters with members of other groups are likely to be restricted to fleeting encounters on campus or in large lecture halls, or on the street in town, thus both quantity and quality of contact with out-group faces should be rather limited.

Therefore, it was assumed that these three groups would not differ with respect to the quality of contact but in the quantity of contact with other persons in general, and with members of other ethnic groups in particular. A considerable number of border patrol/police officers also work in the area of custody pending deportation. As one might expect that this specific group differs with respect to contact quality from the passport control officers, only border patrol officers working at the passport control desks were considered for participation in the present study.

Besides these attempts to investigate factors that are associated with the OEE, it is also worthwhile to compare the recognition performance of police officers with that of bank clerks and students with subgroups of other ethnicity faces. This question is akin to the more general question whether or not police officers are better in recognizing faces than civilians, or more generally, whether or not they are generally better eyewitnesses (see the meta-analysis by Zimmerman and Sporer 2010), a question often addressed to experts before courts of law (R. P. Fisher, personal communication, December 22, 2008).

As border patrol officers are required to interact daily with persons from various cultures and ethnic origins, the job profile demands the ability to differentiate between and within these ethnic groups. We examined whether border patrol officers differed in the recognition performance of faces of different ethnic groups. If increased contact with other ethnic groups leads to a high level of expertise with these faces, an improved recognition of out-group faces could be expected. Hence, border patrol officers should either show no OEE or a reduced OEE.

² Even though we use the terms race and ethnic group, African American and Black, European American and White, interchangeably in most parts of this paper, some researchers have argued that it is the relative status as an out-group that may be responsible for the out-group recognition deficit (Hugenberg et al. 2007).

³ Data for this experiment were collected some while ago when it was still customary to obtain cash from bank clerks rather than an ATM.

The Tasks of Matching Faces and the OEE

To simulate the work tasks of border police officers and bank clerks a delayed-matching task was employed to best capture the cognitive demands inherent in these occupations for recognizing travelers and customers, respectively. In a delayed-matching task, a stimulus is presented for a fixed period of time, followed by a brief mask, which covers the stimulus and is presented for a fixed period of time and prevents further processing of the target stimulus (i.e., face). Thus, the participant has to keep the image of the stimulus in visual working memory (Baddeley 2000) to compare it with the succeeding stimulus face. After the mask, a second stimulus is presented. The participant judges whether or not the first and the second stimulus are identical. Lindsay et al. (1991), using a variation of this task, found an OEE with Black and White participants and Black and White faces. In their study, the White participants performed worse with Black faces, whereas the Black participants performed equally well with both stimulus groups. Although the delayed-matching task is not fully equivalent to the occupational tasks of border patrol officers or bank clerks, it does allow for greater control of face processing by equalizing the matching task between the experimental groups (i.e., border patrol officers, bank clerks, and students).

Sporer et al. (2007) applied a matching task to investigate the OEE with Turkish and Austrian children aged 10–15 years, viewing Turkish and German faces representing people approximately 18–20 years old. The participants had to match ten faces depicted in a three-quarter profile, presented consecutively on separate cards, to the corresponding face on a table with ten faces depicted in a full-face view.⁴ Sporer et al. measured the time it took to correctly match the ten faces and found that the Austrian children were slower to match the Turkish faces compared to the Turkish children. No differences for the German faces were observed. A follow-up study with 64 adolescents replicated these findings, demonstrating that the effect is not restricted to children but also holds for older participants (Sporer, unpublished data, no date). These matching studies suggest that the OEE is found with children and adolescents not only for a recognition task (cf. Pezdek et al. 2003), but that the OEE reflects a more general processing deficit *at the perceptual stage* (Lindsay et al. 1991).

⁴ The change in pose was used to insure that the task involved not just the matching of an identical two-dimensional picture (stimulus matching) but the matching of a face photograph taken from two different angles (see Bruce 1982, regarding the importance of face recognition vs. stimulus recognition).

The IOM as It Applies to Context and Experience

Sporer (2001a) integrated early stages of social-perceptual face processing in his in-group/out-group model (IOM) of face recognition. In particular, Levin's (1996, 2000) evidence for differential categorization times for own- and other ethnicity faces provides the foundation for differential processing of in-group (i.e., own-ethnicity) versus out-group (i.e., other ethnicity) faces when initially encountering a non-familiar face. Since the publication of the IOM, other researchers have presented support for early perceptual categorization processes discriminating between own- and other-group faces as well (e.g., Bernstein et al. 2007; Hugenberg et al. 2007; MacLin and Malpass 2001; Shriver et al. 2009). According to the IOM, faces are perceived and processed differently after initial categorization. In some studies, categorization itself has been sufficient to produce an OEE (e.g., Bernstein et al. 2007). Thus, after initial categorization triggered by out-group specific markers (e.g., MacLin and Malpass 2003), different processing routes may be followed.

According to the IOM, in a recognition task, in-group faces tend to be processed more holistically and automatically, whereas out-group faces are more likely to be processed more feature-based (Sporer 1991). In a matching task (e.g., matching a person to a passport photo), both target and to-be-matched face may be looked at repeatedly, perhaps searching for individual features to be matched (White et al. 2014). These processing differences are not postulated to be all-or-none but relatively more or less dominant depending upon the task and immediate need of an observer. If recognition is required for a task at hand (e.g., as a prerequisite for matching a person to a passport photo), more effort is required for processing. A change in pose, or profile view, of an own- or other-ethnicity face is suggestive of differential processing as postulated by the IOM. Previous research has shown that the OEE is stronger when the pose is changed between presentation and test than when identical stimuli are presented (Argstatter et al. 2002; Meissner and Brigham 2001; Sporer and Horry 2011). Thus, an additional between-subjects factor in the delayed-matching task was implemented to assess the effect of pose change between study and test on own- and other ethnicity faces. One-half of the participants viewed faces in the delayed-matching task depicted only in a full-face pose and one-half viewed test faces depicted in a right, three-quarter profile. If out-group face processing is primarily feature-based, this change in pose of the stimulus face at test was hypothesized to exacerbate the performance deficit with out-group faces but not with in-group faces. Although this manipulation was expected to affect border patrol officers as well, it should do so to a lesser extent compared to bank clerks and students due to officers' experience with more ethnic groups. With in-group faces, which are assumed to be processed mainly holistically, a change in pose at test should not affect recognition as strongly because processing is not expected to rely on a single or a few features of the face.

Hence, the representation of in-group faces is more likely to be invariant to transformations in pose.

Finally, the IOM is the only theoretical account that not only predicts a more general out-group recognition deficit like the own age (Martschuk and Sporer 2018; Wright and Stroud 2002) and own-gender bias (Wright and Sladden 2003) but also a *shift in response criterion* (for a detailed discussion, see Sporer 2001a, b) due to greater perceived out-group homogeneity, of which the OEE may simply be a specific case. Accordingly, we expected a more lenient response criterion for out-group faces compared to in-group faces with bank clerks and students, but not with border patrol officers.

Method

Data were collected to examine the association of amount of contact and occupational context with different ethnic groups on processing of in-group and out-group faces. A recognition test and a delayed-matching task were implemented involving Black, Hispanic, Turkish, and White German faces. To assess the amount of contact with people from different ethnic groups, a questionnaire was administered, which asked for subjective estimates of contact with people from those ethnic groups whose faces were used in the experiments.

Participants

One-hundred-twenty-eight participants ($n = 64$ females) were recruited; 32 were German border patrol officers (Mdn, age = 30) working at the Frankfurt International Airport in Frankfurt, Germany; 32 were German bank clerks (Mdn age = 42) of a small university town in Germany; and 64 were students (Mdn, age = 21) at the University of Giessen, Germany. All participants were White and received sweets as a small reward for participation. Participants were recruited in accordance with ethical guidelines governing human participants in research.

Materials

Faces For the recognition task the stimuli consisted of 160 male faces depicting Blacks ($n = 40$), Hispanics ($n = 40$), Turks ($n = 40$), and Germans ($n = 40$), and 64 different faces of the same ethnic groups for the delayed-matching task (i.e., $n = 16$ faces per stimulus group). The Black and Hispanic faces were provided by Roy S. Malpass and Christian A. Meissner, respectively, and had been successfully employed in OEE studies in the USA with Black, Hispanic, and White participants. The pictures of Turkish and German faces were previously used by Sporer et al. (2007) and Gehrke (2005), and portrayed students at the universities of Giessen and Frankfurt. All of the photographs depicted males between 18

and 30 years old with no distinctive features (e.g., scars, piercings, or ear rings). Faces were depicted in a full-face pose (i.e., frontal view) and a three-quarter profile (i.e., 45° angle, with the right cheek in view) with neutral expressions. Photographs were standardized using Adobe Photoshop, that is, clothing cues were covered so that only the face and the neck were visible, and each photograph measured 432 × 648 pixels.

Contact Questionnaire The questionnaire consisted of items concerning the subjective amount and quality of contact with out-group members. There were three sub-sections adopted from questionnaires used in previous research (see Brigham 1993; Slone et al. 2000; Sporer 1999; Sporer et al. 2007) assessing the quantity, quality, and frequency of contact with other ethnic groups. These questions addressed different types of contact during childhood, work, and school.

In addition, participants estimated the amount of daily contact they experienced with people of other ethnic groups as a function of their occupation (e.g., “If you think about a typical day at work/university, how many persons of the following ethnic origins do you encounter [please estimate by giving an absolute value]?”). The question was followed by a list of geographical regions (e.g., Africa, Near East, and Middle East), for which the participants were requested to give an estimate of daily contact.⁵

Procedure

Different faces were used for the recognition test and the delayed-matching task. The data from the patrol officers were collected directly at the Frankfurt International Airport. Two Macintosh iMac computers were placed in an office of the border patrol department (approximately 3 m × 4 m), so that two participants could complete the experiment at the same time. The data from the bank clerks were collected in a separate room at the bank’s main branch at the Sparkasse Giessen, also in groups of two. Student data were obtained in a laboratory room of the University of Giessen.

Recognition Test The participants sat in front of the computer and completed the instructions presented on the screen. The experimental session began with the recognition test. Participants studied 80 faces, 20 per each ethnic category. Prior to the faces being presented, three different signs (i.e., a white square, white circle, and white cross) were shown

⁵ The application of a questionnaire to ask for subjective data may not be the method of choice as it is subject to distortions regarding the accuracy of frequency estimates as well as due to social desirability considerations. Unfortunately, we were unable to obtain objective data from the institutions involved (there exist no records for the ethnicity of customers at a bank, nor at the airport, nor from independent observations). As preliminary inspections of the distributions of responses to these latter questions revealed that distributions were heavily skewed, we employed only nonparametric statistical analyses with the contact data.

consecutively for 5000 ms on the screen to direct the participants' gaze to the middle of the screen. The faces (20 faces per ethnic group, all in full-face view) were presented consecutively in random order for 5000 ms with an inter-stimulus interval (ISI) of 500 ms. Each picture was centered on the screen. Between the study phase and the test phase, participants were allotted approximately 10 min to complete the childhood portion of the contact questionnaire. During the test phase, the 80 faces from the study phase were randomly intermixed with 80 new faces (20 from each ethnic group). All faces in the test phase were presented consecutively for 5000 ms with an ISI interval of 500 ms in three-quarter profile. Participants indicated whether each face had been seen before using two keys designated for the left and right forefinger at the computer keyboard ("c" and "m").

Delayed-Matching Task After the recognition test, participants were given approximately 10 min to complete the next portion of the contact survey before beginning the delayed-matching task. The delayed-matching task consisted of 64 trials consisting of 16 faces for each ethnic group. At the beginning of the delayed-matching task, participants were given two practice trials to help them get used to the procedure. In every trial, a face was shown in full-face view for 1000 ms. Subsequently, a mask consisting of random dots covering the entire computer screen was presented for 1500 ms followed by a test face that was either the same face or a new face of a different person. The test face was presented either in a full-face view or a three-quarter profile view⁶ for 1000 ms. The matching task always involved a face from the same ethnic group. The participants had to judge, without a time constraint, whether or not the two faces they had seen were of the same person.

After the delayed-matching task, participants were thanked and received sweets as a reward for participation. Border patrol officers and bank clerks were not immediately informed about the purpose of the study until all participants from these groups had completed the experiment. After participation was complete, their supervisors informed them about the specific purpose and provided individual feedback. Similarly, students received feedback after all the data were collected.

Design and Analyses A 3 (participant occupation: border patrol police vs. bank clerks vs. student) by 4 (ethnicity of face: Black vs. Hispanic vs. Turkish vs. White German) mixed design was implemented, with participant occupation as

between-participants and ethnicity of faces as repeated measures factor. The same design was used both for the recognition and delayed-matching data.

The data from the contact questionnaire were analyzed to determine if participants from the three occupational groups indeed showed the postulated differences in contact frequency. The results from the recognition test and the delayed-matching task are reported using the signal detection theory measures A' as an estimate of recognition performance (Rae 1976) and B'' as an estimate of response bias (Donaldson 1992, 1993) as dependent variables. Recognition performance scores range from 0 (no recognition), 0.5 (chance level performance), to 1 (perfect recognition), while response bias scores range from -1 (liberal response bias, i.e., higher likelihood of saying "yes" to a face) to 1 (conservative response bias, i.e., lower likelihood of saying "yes" to a face). Effect sizes are reported as partial η^2 for significance tests of the F distribution with $df > 1$ in the numerator as well as for any interactions. For pairwise comparisons, effect sizes are reported as Cohen's d based on the means and standard deviations for the respective cells. Note that for repeated measures, the formula for d differs from that for the between-participants comparisons in so far as the correlation between the dependent measures has to be considered (for a mathematical proof, see Dunlap et al. 1996; see also Borenstein 2009). This point is crucial. Past studies may or may not have calculated d using these correlations (at least they do not report the correlations between in-group and out-group performance measures; see Sporer and Cohn 2011, for further discussion of this point).

Results

Contact Questionnaire

There were three types of questions in the contact questionnaire—general questions about contact with members of other ethnic groups as used in many previous studies, questions about estimated relative amount of contact, and several specific questions regarding the estimated frequencies of daily contact with Blacks from Africa, people from the Near East, Middle East, Far East, Non-German Whites, Whites, and Germans. Table 1 depicts the medians of estimated daily contact for border patrol police, bank clerks, and students. Separate nonparametric Kruskal-Wallis analyses of variance with participant occupation revealed significant differences in the frequencies of contact as a function of occupation for all questions, all $\chi^2(2, N = 117) > 47.78$, all $ps < .001$.

Patrol officers indicated substantially more contact with all groups (i.e., own- and other-ethnic groups) compared to bank clerks and students (see Table 1). Perhaps unexpectedly, students had somewhat more contact than bank clerks, in particular, with Germans. Bank clerks, alternatively, had the largest

⁶ Originally, for half of the students and half of the border patrol officers, the test face had been presented in a full-face view, for the other half in three-quarter view. As the results of the participants presented with faces twice in frontal view were close to ceiling effects, these data were dropped from any inferential analyses as suggested by one of the reviewers (Dan B. Wright). As the data from the bank clerks were collected later, they were all tested with a change in pose between first and second stimulus presentation.

Table 1 Medians of estimated daily contacts with different ethnic groups

Contact group	Students	Bank clerks	Patrol officers
Blacks	3.0	0.0	80.0
Near East	6.5	3.0	100.0
Middle East	2.0	0.0	65.0
Far East	2.5	0.0	275.0
Non-German Whites	2.0	2.0	200.0
Whites	5.0	2.0	300.0
Germans	80.0	30.0	400.0
Years of Contact	1.0	17.0	7.0

All differences between patrol officers and the other two groups were significant at $p < .001$ (Kruskal-Wallis ANOVAs)

median numbers of years of contact (as a function of their higher age) in comparison to the two other groups. As expected border patrol officers had more experience with other ethnic groups providing a substantive foundation upon which to test whether the quantity of other-group contact may be associated with the OEE.

Recognition

OEE and A' A 3 (participant occupation) by 4 (ethnicity of face) mixed-model analysis of variance (ANOVA) was conducted with A' as the dependent variable. There was a main effect for ethnicity of face, $F(3,375) = 27.73, p < .001, MSE = .01, partial \eta^2 = .182$. Follow-up analyses revealed that German faces were recognized best ($A' = .698, CIs = .675, .721$), followed by Turkish ($A' = .674, CIs = .651, .697$), Hispanic ($A' = .623, CIs = .600, .646$), and Black faces ($A' = .563, CIs = .538, .588$). In particular, the comparisons between recognition of German faces and out-group faces were as expected. There was a large OEE comparing German with Black faces, $t(127) = 9.30, p < .001, d = 0.99$; a medium OEE with Hispanic faces, $t(127) = 5.92, p < .001, d = 0.58$; and a small OEE with Turkish faces, $t(127) = 1.99, p = .049, d = 0.19$. On average, in-group faces were recognized better than all three types of out-group faces, $t(127) = 7.50, p < .001$; that is, the overall OEE was $d = 0.67$ (CIs 0.481; 0.861).

There was no main effect for type of occupation, $F(2,125) = 2.23, p = .112, MSE = .04, partial \eta^2 = .034$. Planned contrasts indicated that recognition performance of patrol officers ($A' = .663, CIs = .630, .696$) was higher than of bank clerks ($A' = .613, CIs = .581, .646$), $F(1,125) = 4.51, p = .036, d = 0.53$, but not higher than that of students ($A' = .641, CIs = .617, .664$), $F(1,125) = 1.16, p = .283, d = 0.28$.

The expected interaction between participant occupation and type of face is depicted in Fig. 1, $F(6,375) = 2.74, p = .013, MSE = .01, partial \eta^2 = .042$. Follow-up analyses with

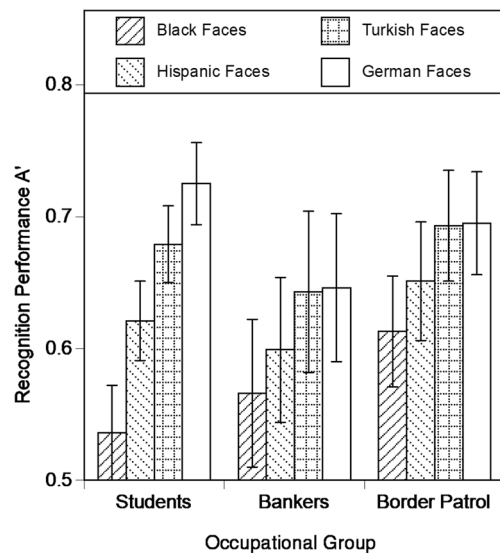


Fig. 1 Means (and 95% CIs) for recognition performance A' of students, bank employees, and border patrol officers for Black, Hispanic, Turkish, and German faces

Bonferroni-adjusted comparisons ($p < .05$) revealed that patrol officers were better with Black faces than students, while students were better with German faces than bank clerks. There were no other significant differences.

OEE and response bias (B'') A 3 (participant occupation) by 4 (ethnicity of face) ANOVA with response bias B'' revealed a main effect for ethnicity of face, $F(3,375) = 9.98, p < .001, MSE = .15, partial \eta^2 = .074$; all other effects were nonsignificant ($F_s < 1.76$). Participants were significantly more conservative with Black ($M = .295, CIs = .204, .386$) and German faces ($M = .269, CIs = .183, .355$) than with Hispanic ($M = .050, CIs = -.044, .144$) and Turkish faces ($M = .085, CIs = -.006, .176$).

Delayed-Matching and OEE

Analogous to the recognition task, sensitivity (A') and bias (B'') were calculated for participants' responses to the delayed-matching task. Hits were recorded when participants responded "identical" and the same face was depicted in both the view and test photographs. Responding "identical" to a different person at test was coded as a false alarm. Recall that the data utilized for the following analyses involved only participants who viewed faces depicted in full-face during viewing and a three-quarter profile during test. After having collected data for the patrol officers and students, preliminary analyses had revealed a ceiling effect when faces were shown in identical frontal views, that is, all group A' scores $> .97$, so we did not test these differences for significance due to these ceiling effects.

From an applied perspective, the error rates (i.e., false alarms) are still of interest. With no change in pose, there were .020 errors with Black faces, .047 with Hispanic, .023 with Turkish, and .004 with German faces, respectively. However,

errors with a change in pose were remarkably higher (Black faces: .147; Hispanic faces: .145; Turkish faces: .055; German faces: .030). In other words, Blacks and Hispanics would be much more frequently falsely considered as the person on the ID card than Turks and Germans.

As the experiment with bank clerks was carried out after collecting data with patrol officers and students, we decided to test bank clerks and a second group of students only with a change in pose between study and test from frontal to three-quarter view. Hence, analyses were carried out with a reduced sample of 48 students, 32 bank clerks, and 16 patrol officers who had all been tested in the change of pose condition.

OEE and A' A 3 (participant occupation) by 4 (ethnicity of face) mixed-model ANOVA parallel to the recognition data above was conducted on A' for the delayed-matching task. There was no main effect of participant occupation, $F(2,93) = 1.02, p = .366, MSE = .01, \text{partial } \eta^2 = .021$. Patrol officers ($A' = .951; CIs: .929, .973$) were not significantly better than bank clerks ($A' = .940; CIs: .925, .956$) and students ($A' = .933; CIs: .920, .946$; planned contrast for police vs. non-police: $F(1,93) = 2.59, p = .111, d = 0.33$).

There was a main effect of face ethnicity, $F(3,279) = 33.44, p < .001, MSE = .00, \text{partial } \eta^2 = .264$. Participants matched German faces ($A' = .968, CIs = .956, .980$) better than Black ($A' = .889, CIs = .871, .907$), $t(95) = 8.09, p < .001, d = 1.05$, and Hispanic faces ($A' = .936, CIs = .924, .947$), $t(95) = 5.41, p < .001, d = 0.57$, but not better than Turkish faces ($A' = .961, CIs = .953, .969$), $t(95) = 1.17, p = .244, d = 0.14$. Parallel to the recognition task, in-group faces were matched better than all three types of out-group faces, $t(127) = 7.50, p < .001$; thus, the overall OEE was $d = 0.636 (CIs: .446; .827)$. The interaction between participant occupation and face ethnicity was not significant, $F(6,279) = .21, p = .973, MSE = .00, \text{partial } \eta^2 = .005$ (see Fig. 2).

OEE and B'' As in the recognition task, there was a significant main effect of face ethnicity on response bias, $F(3,279) = 13.19, p < .001, \text{partial } \eta^2 = .124$. Participants were more likely to indicate “identical” (i.e., demonstrate a liberal bias) for Hispanic faces ($B'' = -.314, CIs = -.441, -.187$) compared to Turkish ($B'' = .051, CIs = -.066, .168$), Black ($B'' = .160, CIs: .024, .296$), and German faces ($B'' = .204, CIs: .103, .306$). There was no main effect for participant occupation and no interaction effect, both $F_s < 1$.

Gender and Age Effects

Table 2 (at the bottom) depicts nonparametric Spearman-Brown correlations between gender and age of participants with performance A' for recognizing out-group faces for the three participant groups. In the recognition task, positive correlations indicated that females performed slightly, but

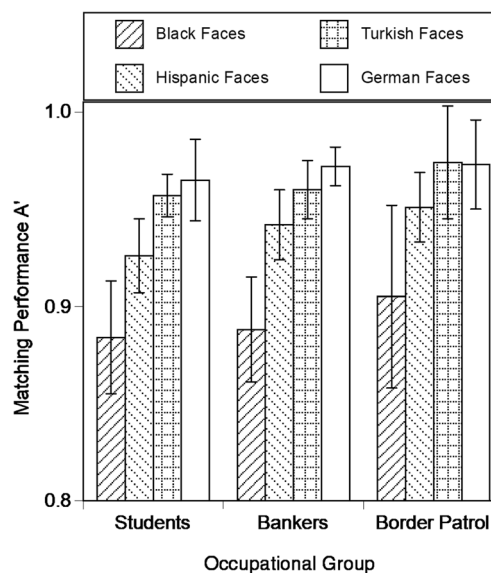


Fig. 2 Means (and 95% CIs) for matching performance A' of students, bank employees, and border patrol officers for Black, Hispanic, Turkish, and German faces

nonsignificantly better than males (mean, $\rho = .22$), while at delayed-matching both genders performed fairly equally (mean, $\rho = -.05$).

Perhaps of interest with respect to developmental trends in face recognition ability, the correlations between participant age with the performance measure A' in the recognition task averaged across the three out-group face groups showed overall nonsignificant negative correlations for students ($-.16$), bank employees ($-.29$), and patrol officers ($-.37$). Similarly, nonsignificant negative correlations with age were observed with the matching task ($-.11, -.24, -.50$). These results are in line with recent findings that a decline in face recognition ability may set in earlier not just at advanced old age (see Martschuk and Sporer 2018).

OEE, Contact, and Face Recognition

For each of the three groups, separate nonparametric correlational analyses were conducted with participant responses to the contact questionnaire and out-group recognition performance (A'). Out-group recognition performance was calculated as mean A' of the three out-group face groups (i.e., Black, Hispanic, and Turkish faces) for the recognition and the delayed-matching task. Table 2 displays the Spearman rank correlations between contact scores and mean A' for border patrol officers, bank clerks, and students. Although some of the correlations were significant in the predicted direction, the number of significant correlations was smaller than what one would have expected by chance. Results did not differ when looking at correlations for stimulus ethnicities of faces separately.

Table 2 Nonparametric correlations of quantity and quality of contact questions with performance A' in the recognition and delayed-matching task

Contact questions	Recognition task			Delayed-matching task		
	Students (n = 64)	Bank (n = 32)	Police (n = 32)	Students (n = 48)	Bank (n = 32)	Police (n = 16)
General contact questions						
% children in grammar school	.19	.11	.38	.23	.16	-.03
% children in middle school	.04	.16	.32	.15	-.03	-.17
% in high/vocational school	-.03	.23	.05	-.04	.19	.36
% children in neighborhood	.04	-.15	.34	.31	-.15	.11
# friends in grammar school	-.06	-.07	.12	-.06	-.24	.14
# friends in middle school	.08	.23	.11	.09	.09	.13
# friends in high/vocational school	.16	.09	.14	.17	.07	.44
> 2 contacts per week at university/training	-.05	.15	.24	.18	.22	--
> 2 contacts per week at leisure activities	-.03	.35	.21	-.02	-.03	-.12
> 2 contacts per week at work	.00	.08	-.26	.24	-.06	-.08
> 2 contacts per week in shops	.14	.15	-.18	.09	.03	-.23
> 2 contacts per week in neighborhood	.11	-.20	.06	-.03	-.42	.03
# known by nickname	.10	.17	.08	.17	-.08	-.14
# close friends	.02	-.05	.01	.08	.27	.13
# going on a date	.07	.07	.14	.03	-.32	.21
# years in Germany	-.12	-.29	-.39	-.23	-.24	-.50
Relative experience with Blacks at work	.16	.23	-.19	.37*	-.30	.15
Relative experience with Hispanics at work	-.17	.05	-.16	-.02	-.35	-.21
Relative experience with Turks at work	.10	.11	-.06	-.04	-.10	.29
Relative experience with Blacks in media	.16	.15	.08	.04	-.25	.52
Relative experience with Hispanics in media	-.08	.22	.14	-.04	-.30	.34
Relative experience with Turks in media	.08	.04	.16	.13	-.03	.34
Frequency estimates						
Daily frequency of contact with Black Africans	.06	-.00	.02	.19	-.09	.13
Daily frequency of contact with Near East	.26	.03	.07	.17	-.07	.31
Daily frequency of contact with Middle East	.24	-.06	-.11	.21	-.18	.36
Daily frequency of contact with Far East	.25	.07	.05	.11	.02	.51
# years contact with out-groups	-.15	-.23	-.01	.13	-.24	-.26
Mean correlations	.06	.06	.05	.10	-.09	.11
Gender of participants	.10	.40	.15	-.08	.24	-.31
Age of participants	-.16	-.29	-.37	-.11	-.24	-.50

For the frequency questions, 9 participants were dropped due to an incorrect answer format

* $p < .05$ (unadjusted)

As the numerous nonsignificant correlations in Table 2 show, self-reported contact with ethnic out-groups does not seem to be associated with better recognition performance. Perhaps, the only exception to this pattern of null findings was a significant positive correlation (.42) with patrol officers between the number of years on the job and recognition performance A' for Black faces.

At the end of the questionnaire, participants had also been asked to estimate their performance in the subsequent recognition task and some additional questions about job experience. Only the patrol officers could somewhat gauge their own performance, with $r(30) = .57, p < .01$, for Black faces, $r(30) = .51, p < .01$, for Hispanic faces, and $r(30) = .22$ and $.14$, both ns, for Turkish and German faces, respectively. Perhaps, the only

significant correlation worth noting was with patrol officers between the number of years on the job and recognition performance A' for Black faces, $r(30) = .42, p < .05$.

Discussion

The main goal of this study was to test if job-related quantity of contact with persons of different ethnic groups results in better recognition and face matching performance with faces from these groups. This goal was accomplished by testing three distinct occupational groups in Germany, border patrol officers, bank employees, and students, who were postulated to differ in quantity of out-group contact. We assumed that performance with out-group faces of African Americans, Hispanics, and Turks would be worse compared to in-group faces of Germans, and that recognition performance would increase with the likelihood of contact with the respective out-groups. As contact should be least likely with Blacks and Hispanics from the USA, the processing deficits for these faces should be stronger than for faces of Turks which make up the largest minority, approximately 3%, of the German population.

Experience and Out-Group Member Processing

We demonstrated that participant occupation was associated with differences in the frequency of daily contact with out-group members, which in turn was associated with differences in recognition and matching performance. Border patrol officers had more contact with out-group people than bank clerks and students. As predicted, for the standard recognition and delayed-matching tasks, Black faces were ostensibly processed the worst, followed by Hispanic, Turkish, and German faces as evidenced by the effect sizes for the OEE, with large effects for Black faces ($d = 0.98$ for standard recognition and 1.05 for delayed-matching), medium OEE sizes for Hispanic faces ($d = 0.57$ for standard recognition and 0.57 for delayed-matching), and small OEE sizes for Turkish faces ($d = 0.19$ for standard recognition and 0.14 for delayed-matching). This pattern follows the likely amount of contact participants have with the respective out-groups (see Table 1). These results support the contact hypothesis in terms of mere frequency of encounter (e.g., Tredoux et al. 2020; Valentine et al. 1995).

Also supporting the notion of the frequency of encounter was the interaction effect obtained in the recognition task between participant occupation and face ethnicity. Patrol officers were relatively better with Black faces compared to students who were better with own-group, German faces. Thus, the amount of daily professional contact may be associated with an attenuation of the OEE in some instances. This is noteworthy as the effect between participant group and ethnicity of

faces was primarily driven by Black faces that showed the strongest recognition deficit. However, the relative improvement with patrol officers was small and restricted to Black faces. Despite a similar pattern, these effects did not replicate with the delayed-matching task, presumably due to the low power after deleting half the sample because of a floor effect. Perhaps, as with Asian faces, which appear particularly difficult to process for Europeans (e.g., Rhodes et al. 1989), increased amounts of exposure may be associated with an attenuation of OEE effects. For example, Dixon (1994; cited in Valentine et al. 1995) found that British interns improved in their recognition ability of Black faces after 8 weeks staying abroad in a Black African community compared to a control group not traveling abroad.

Methodological Considerations

Measuring Out-Group Contact It is also noteworthy that the self-reported amount of contact within the three groups was not related to the performance as measured with the two tasks. Note, however, that these correlations were only calculated within a single ethnic group (Germans). Perhaps, previous studies may have found some positive relationships when calculating correlations *across* groups. The null findings observed here may also be due to the questionnaire used. People may simply not be very good with estimating frequencies of contact, which varied extremely from participant to participant in each group. Regarding the other questions, which have been widely used by other authors in previous studies, the lack of predictive power is not surprising as these types of questions have never explained much of the variance in recognition performance (less than 2% of the variance according to the meta-analysis of Meissner and Brigham 2001). Consequently, asking people about the amount of past and daily experience with out-group faces may not be a reliable, and hence also not a valid way to post-dict their face recognition ability. This may also have important implications when questioning witnesses in court, trying to assess whether they are likely to be correct when they had identified a person from another ethnic group. If witnesses cannot reliably estimate their contact, and if such estimates are not related to recognition performance, answers to such questions should not be used to assess whether a witness may be good or bad in recognizing out-group faces, let alone whether he or she is correct with respect to a specific identification.

Face Processing of Multiple Out-Groups This is one of the few studies in the literature that used large sets of natural faces from four different ethnic groups that differed in the view depicted (i.e., full-face to three-quarter profile and vice versa) between study and test. All faces had been pre-rated for distinctiveness, memorability, and other aspects important for

recognition in previous studies which had used both in-group and out-group members in fully crossed designs, demonstrating consistent OEE effects. Thus, the results cannot be attributed to idiosyncrasies of the stimulus faces. Our methodology also fulfills the requirement of stimulus sampling (Wells and Windschitl 1999) that postulates that experimental studies ought to utilize large samples of stimuli, not only large samples of participants to obtain ecologically valid results. This being said, we acknowledge that it would have been desirable to also use female faces. We do emphasize, however, that equal numbers of male and female participants were randomly assigned to conditions in all three groups. However, adding female faces would have also increased the complexity of the design, requiring a much larger number of participants. Also, at the time, no pre-rated stimulus sets of female faces were available.⁷

Other-group-based characteristics (i.e., age and gender) are also important to consider in relation to face processing. Past meta-analyses concluded that both children (e.g., < 13 years) or older adults (in the literature usually defined as > 60 years) showed worse recognition performance as well as own-age biases regarding faces of their own-age group (Martschuk and Sporer 2018; Rhodes and Anastasi 2012). However, given the age of participants in the present study, no such effects were expected nor observed. Future studies should consider gender and age of the participants, and gender and age of the stimulus faces in relation to experience and potential interaction effects on face processing and recognition. Although both superior performance of female participants (Shapiro & Penrod, 1986) as well as an own-gender advantage (Herlitz and Lovén 2013) may be expected, there is no theoretical reason to assume that these factors may interact with ethnicity effects. With female faces, the particular importance of hair and hairstyle will need to be considered (Wright and Sladden 2003).⁸

Applied Implications

From an applied perspective, our research is one of a small set of studies comparing the performance of police officers, in this study border patrol officers, with that of civilians (for a meta-analysis, see Zimmerman and Sporer 2010). Here, patrol officers were somewhat better in the recognition ($d = 0.38$) and in the delayed-matching task ($d = 0.33$) but both effects failed to reach significance ($ps = .063$ and $.111$, respectively). However, at least with Black faces, the border patrol officers

were superior to the two other groups in recognition performance, presumably due to their increased contact.

In the delayed-matching task, performance was extremely good when stimulus faces between study and test were identical but much worse when face pose changed. The OEE observed with the delayed-matching task when pose was changed ($d = 0.64$) was comparable in magnitude to the effect size for the recognition task ($d = 0.67$) reported above. For any type of security personnel, this type of change as well as any other types of changes in appearance (e.g., change in age, hairstyle, glasses, or emotional expression) is likely to be the norm and may pose serious problems (see also Kemp et al. 1997). Although the false alarm rates observed here may sound small in an absolute sense (between 5 and 15% for out-group compared to 3% for in-group faces), they must be considered a problem when maximum security issues are at stake.

At a more general level, we should also note that identifying or matching passers-by, or their images, with photographs or video material appears to be a difficult task not only with faces of other ethnic groups but in general (see Henderson et al. 2001; Kemp et al. 1997). Some of the difficulties may stem from the poor qualities of photographs used, or from changes in pose or appearance between the person to be recognized and the display, although human performance appears to be rather robust in light of such changes (Liu et al. 2003). Perhaps, portraying persons from multiple perspectives (e.g., full-face and three-quarter views) on passports may lead to improvements in matching performance. Whether “change in appearance” instructions (see Charman and Wells 2007) would be useful in this context remains an empirical issue that would have to be tested with a paradigm more akin to the present study.

A related criminal justice problem the present studies address only indirectly concerns identifications based on videotapes from surveillance cameras (see Davies and Thasen 2000; Davis and Valentine 2009), computerized biometric face matching and matching performance of forensic pathologists and forensic anthropologists. These professionals are often called as experts to evaluate the identity of faces depicted on videos or photos and faces of suspects (or faces of deceased). Our research bears on the evaluation of these investigations because signal detection theory provides a statistical framework with which performance in these domains can be assessed. For example, in biometric matching, the performance of the algorithms used are usually only evaluated with reference to hits, thus ignoring the equally important question how often false alarms occur when a face is wrongly identified as being identical to the image at hand. In the psychological literature, recognition of faces, and probably matching as well, are usually considered to involve holistic processes that occur rather automatically, not feature-based analytically, at least for in-group faces. By comparison, out-group faces may be analyzed more analytically, comparing individual features (see Sporer 2001a). Interestingly, forensic anthropologists, at least in Germany, usually proceed analytically, going through a long list

⁷ As part of the grant application of the German psychological association, the first author had to demonstrate that the design had sufficient power to obtain statistically significant results. However, it was not foreseeable that half of the participants in the same-pose conditions would yield ceiling effects and therefore had to be dropped.

⁸ The importance of hair, for both male and female faces, casts doubts on studies that have only used the inner parts of a face or of artificially created faces as often used in the neuropsychological literature to gain optimal control over stimuli.

of features one-by-one, tallying the number of specific feature correspondences between a face to be identified (e.g., from a surveillance video camera or speed camera) and a photo of a suspect (Holley 2012; Rösing 1999). To our knowledge, no systematic research exists to evaluate the accuracy rates of these assessments within the framework of signal detection theory advocated here.

Open Questions, Limitations, and Concluding Comments

The demonstration of an OEE with a delayed-matching task implies that the OEE is not only a memory phenomenon but seems to be located at earlier stages of encoding (cf. Lindsay et al. 1991; Sporer 2001a). In another study, we have also noted that the OEE of German participants with Turkish faces was much stronger when the pose was changed between study and test ($d = 0.41$) than when the pose remained the same and where there was no significant OEE ($d = 0.08$; Argstatter et al. 2002; Sporer and Horry 2011). The importance of change in pose also supports the notion that recognizing or matching a face involves more than mere stimulus recognition, requiring the abstraction of a face code from a stimulus photograph (Bruce 1982). Hence, studies using identical stimuli at presentation and test may not be adequate for studying out-group processing deficits in both standard recognition and matching paradigms.

The quasi-experimental design of this study does not allow one to test rival theoretical accounts (see also Wright 2006). Thus, the data are compatible with Valentine's (1991) multi-dimensional face space model, Hugenberg et al.'s (2007) work on social categorization, and Sporer's (2001a) IOM, although only the latter also predicts a response bias. Of note, and as predicted by the IOM, several patrol officers informally stated that they had automatically tried to categorize the faces. The officers tried to assess the originating countries of the faces they encountered. This process is probably highly over-learned because it is a fundamental aspect of the task to find out if the holder of a passport is a citizen of the country that issued the passport.⁹ One officer commented that quite often Chinese citizens try to enter Europe on a Japanese passport. When we presented the experimental procedure to the officers in charge at the airport, they commented that they were spontaneously trying to guess if the Turkish faces came from Morocco or Tunisia. Concerning airport security, it might be valuable to examine the ability of border police officers to correctly categorize persons with respect to their national origin based on a target's ethnic appearance. Congruent with Sporer et al. (2007), the current findings demonstrate that

the ability of border police officers to match faces is affected by the OEE, with worse performance for out-group faces.

The current findings, however, do not address the specific aspects of a face that may have been attended to that allowed patrol officers to be relatively better with Black faces and students better with White German faces. If security personnel are to be trained to improve performance in either recognition or matching tasks, special encoding instructions, perhaps like those used by Hills and Lewis (2006), might be useful. The usefulness of such attention focus instructions should be accompanied by eye tracking methodology that directly determines those parts of faces people attend to when processing and encoding in- and out-group faces.

However, research by Caldara and colleagues has shown that eye movements under natural viewing conditions do not provide conclusive answers about information use. One way to overcome this limitation is to use gaze-contingent techniques which can have a better control of the information feeding the visual system. Such approaches have been successfully applied in the investigation of foveal information extraction from faces and extrafoveal information from scenes (e.g., Caldara et al. 2010; Mielle et al. 2010). There may also be inter-individual differences in face recognition and face matching which have only recently become the focus of attention (see the research on "super-recognizers"; e.g., Bobak et al. 2016; Davis et al. 2016). Whether or not these approaches lend themselves to screening of suitable personnel at airports or other security-systems is an interesting question for the future

In conclusion, the evidence here suggests that border patrol officers' recognition performance benefits from exposure to multiple ethnic groups. Although a performance gap persisted, the OEE was attenuated for these officers. The consequences of error, however, are more serious when it comes to identifying criminals in police investigations, terrorists at the airport, or other security-sensitive situations. Thus, the implication is that security or police work that involves the identification (or photo matching) of diverse groups of people would be facilitated by a diverse work force. Essentially, employing members of different major ethnic groups in jobs where people have to be screened at airports, border crossings, etc. could minimize the probability of identification errors due to OEE. Of course, this can be achieved much easier in countries with a population from diverse ethnic backgrounds such as the USA compared to countries with less ethnic variation such as Germany. As the notion of the "global citizen" continues to manifest itself, it would be a pro-active stance for government agencies and security organizations to diversify the ethnic groups employed to match a location. Moreover, training that informs officers of the perceptual challenges inherent in processing out-group faces in conjunction with increased exposure may further attenuate the OEE and enhance recognition.

⁹ Perhaps, in future studies with airport patrol officers, groups of officers who deal with international customers could be compared with officers responsible for domestic flights.

Acknowledgments Portions of the data were collected as parts of the doctoral thesis by the second author. We are grateful to Annegret Braun, Sakine Canerik, Gholam Yeganeh, and Erdinc Dogan for their help in taking the photographs of Turkish stimulus persons, and to Roy S. Malpass for providing us with photographs of African American, and Christian A. Meissner for photographs of Hispanic and European American faces. We also thank Erdinc Dogan, Andrea Schergung, and Eva Schäfer for collecting portions of the data. We also acknowledge the comments by Colin G. Tredoux and Dan B. Wright on earlier drafts of this manuscript.

Funding Information Open access funding provided by Projekt DEAL. This research was supported by a grant from the Deutsche Forschungsgemeinschaft (Sporer 262/9-1 and Sporer 262/9-2) to the first author.

Compliance with Ethical Standards

Ethical Approval All procedures performed in this study were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Argstatter H, Gehrke J, Hueweler A, Kampschulte A, Schraub EM, Sporer SL (2002) Wiedererkennen von Eigen- und Fremdgruppen-Gesichtern [Recognizing in-group and out-group faces]. Paper presented at the 44th Tagung experimentell arbeitender Psychologen in Chemnitz
- Baddeley A (2000) Short-term and working memory. In: Tulving E, Craik FIM (eds) *The Oxford handbook of memory*. Oxford University Press, Oxford/New York, pp 77–92
- Bernstein MJ, Young SG, Hugenberg K (2007) The cross-category effect: mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychol Sci* 18:706–712
- Bobak AK, Hancock PJ, Bate S (2016) Super-recognizers in action: evidence from face matching and face memory tasks. *Appl Cogn Psychol* 30:81–91
- Borenstein M (2009) Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC (eds) *The handbook of research synthesis and meta-analysis*, 2nd edn. Russell Sage, New York, pp 221–235
- Bothwell RK, Brigham JC, Malpass RS (1989) Cross-racial identification. *Personal Soc Psychol Bull* 15:19–25
- Brigham JC (1993) College students' racial attitudes. *J Appl Soc Psychol* 23:1933–1967
- Bruce V (1982) Changing faces: visual and non-visual coding processes in face recognition. *Br J Psychol* 73:105–116
- Caldara R, Zhou X, Miellet S (2010) Putting culture under the 'Spotlight' reveals universal information use for face recognition. *PLoS One* 5(3):e9708
- Chance JE, Goldstein AG, McBride L (1975) Differential experience and recognition memory for faces. *J Soc Psychol* 97:243–253
- Charman SD, Wells GL (2007) Eyewitness lineups: is the appearance-change instruction a good idea? *Law Hum Behav* 31:3–22
- Chiroro P, Valentine T (1995) An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly J Experimental Psychol Sect A: Human Exp Psychol* 48:879–894
- Chiroro PM, Tredoux CG, Radaelli S, Meissner CA (2008) Recognizing faces across continents: the effect of within-race variations on the own-race bias in face recognition. *Psychon Bull Rev* 15:1089–1092
- Davies G, Thasen S (2000) Closed-circuit television: how effective an identification aid? *Br J Psychol* 91:411–426. <https://doi.org/10.1348/000712600161907>
- Davis JP, Valentine T (2009) CCTV on trial: Matching video images with the defendant in the dock. *Appl Cogn Psychol* 23:482–5005. <https://doi.org/10.1002/acp.1490>
- Davis JP, Lander K, Evans R, Jansari A (2016) Investigating predictors of superior face recognition ability in police super-recognizers. *Appl Cogn Psychol* 30:827–840. <https://doi.org/10.1002/acp.3260>
- Dixon, R. (1994) A prospective investigation of the contact hypothesis as an explanation of the own-race bias in face recognition. (Unpublished undergraduate dissertation, Department of Psychology, University of Durham, Durham, England)
- Donaldson W (1992) Measuring recognition memory. *J Exp Psychol* 121:275–277
- Donaldson W (1993) Accuracy of d' and A' as estimates of sensitivity. *Bull Psychon Soc* 31:271–274
- Dunlap WP, Cortina J, Vaslow JB, Burke MJ (1996) Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods* 1:170–177
- Dunning D, Li J, Malpass RS (1998) Basketball fandom and cross-race identification among European Americans: another look at the hypothesis. Paper presented at the biennial conference of the American Psychology-Law Society in Redondo Beach, CA
- Feingold GA (1914) The influence of context on the identification of persons and things. *J Criminal Law Criminol* 5:39–51
- Garrett BL (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge
- Gehrke, J. (2005). "Die sehen ja alle gleich aus". Einflussfaktoren der unterschiedlichen Wiedererkennensleistung von Gesichtern der eigenen Ethnie und Gesichtern anderer Ethnien (cross-race bias) ["They all look alike to me." Determinants of differential recognition of faces of one's own ethnic group and faces of other ethnic groups (cross-race bias)]. Unpublished Doctoral Dissertation, University of Giessen, Giessen, Germany.
- Goldstein AG, Chance JE (1976) Measuring psychological similarity of faces. *Bull Psychon Soc* 7:407–408
- Goldstein AG, Chance JE (1980) Memory for faces and schema theory. *J Psychol* 105:47–59
- Henderson Z, Bruce V, Burton AM (2001) Matching the faces of robbers captured on video. *Appl Cogn Psychol* 15:445–464
- Herlitz A, Lovén J (2013) Sex differences and the own-gender bias in face recognition: a meta-analytic review. *Vis Cogn* 21:1306–1336. <https://doi.org/10.1080/13506285.2013.823140>
- Hills PJ, Lewis MB (2006) Reducing the own-race bias in face recognition by shifting attention. *Q J Exp Psychol* 59:996–1002
- Holley S (2012) Bildidentifikation: Bild- und Videodokumente als Grundlage der Personenidentifikation [Image identification: photo and video documents as a means for person identification]. In:

- Haffner H-T, Skopp G, Graw M (eds) *Begutachtung im Verkehrsrecht*. Springer-Verlag, Heidelberg, pp 327–366
- Horry R, Wright DB (2009) Anxiety and terrorism: automatic stereotypes affect visual attention and recognition memory for White and Middle Eastern faces. *Appl Cogn Psychol* 23:345–357
- Hugenberg K, Miller J, Claypool HM (2007) Categorization and individuation in the cross-race recognition deficit: toward a solution to an insidious problem. *J Exp Soc Psychol* 43:334–340
- Kemp R, Towell N, Pike G (1997) When seeing should be believing. *Appl Cogn Psychol* 11:211–222
- Levin DT (1996) Classifying faces by race: the structure of face categories. *J Exp Psychol Learn Mem Cogn* 22:1364–1382
- Levin DT (2000) Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition. *J Exp Psychol-Gen* 129:559–574
- Lindsay DS, Jack PC Jr, Christian MA (1991) Other-race face perception. *J Appl Psychol* 76:587–589
- Liu CH, Seetzen H, Burton AM, Chaudhuri A (2003) Face recognition is robust with incongruent image resolution: relationship to security video images. *J Exp Psychol Appl* 9:33–41
- MacLin OH, Malpass RS (2001) Racial categorization of faces: the ambiguous race face effect. *Psychol Public Policy Law* 7:98–118
- MacLin OH, Malpass RS (2003) The ambiguous-race face illusion. *Perception* 32:249–252
- Malpass RS (1990) An excursion into utilitarian analysis. *Behav Sci Res* 24:1–15
- Malpass RS, Kravitz J (1969) Recognition for faces of own and other races. *J Pers Soc Psychol* 13:333–334
- Martschuk N, Sporer SL (2018) Memory of faces in old age: a meta-analysis. *Psychol Aging* 33:904–923. <https://doi.org/10.1037/pag0000282>
- McKone E et al (2019) A critical period for faces: other-race face recognition is improved by childhood but not adult social contact. *Nat Sci Rep* 9:1–13
- Meissner CA, Brigham JC (2001) Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol Public Policy Law* 7:3–35
- Miellet S, Zhou X, He L, Rodger H, Caldara R (2010) Investigating cultural diversity for extrafoveal information use in visual scenes. *J Vis* 10:1–18
- Ng W-J, Lindsay RCL (1994) Cross-race facial recognition: failure of the contact hypothesis. *J Cross-Cult Psychol* 25:217–232
- Pezdek K, Blandon-Gitlin I, Moore C (2003) Children's face recognition memory: more evidence for the cross-race effect. *J Appl Psychol* 88:760–763
- Platz SJ, Hosch HM (1988) Cross-racial ethnic eyewitness identification: a field-study. *J Appl Soc Psychol* 18:972–984
- Rae G (1976) Table of A'. *Percept Mot Skills* 42:98
- Rhodes MG, Anastasi JS (2012) The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychol Bull* 138:146–174. <https://doi.org/10.1037/a0025750>
- Rhodes G, Brake S, Taylor K, Tan S (1989) Expertise and configural coding in face recognition. *Br J Psychol* 80:313–331
- Rösing FW (1999) Standards für die anthropologische Identifikation lebender Personen auf Grund von Bilddokumenten [Standards for the anthropological identification of living persons on the basis of photographic documents]. *Rechtsmedizin* 9:152–154. <https://doi.org/10.1007/s001940050100>
- Shapiro PN, Penrod, S (1986) A meta-analysis of facial identification studies. *Psychol Bull* 100, 139–156
- Shriver ER, Young SG, Hugenberg K, Bernstein MJ, Lanter JR (2009) Class, race, and the face: social context modulates the cross-race effect in face recognition. *Personal Soc Psychol Bull* 34:260–274
- Slone AE, Brigham JC, Meissner CA (2000) Social and cognitive factors affecting the own-race bias in whites. *Basic Appl Soc Psychol* 22:71–84
- Sporer SL (1991) Deep–deeper–deepest? Encoding strategies and the recognition of human faces. *J Exp Psychol Learn Mem Cogn* 17:323–333
- Sporer SL (1999) The own-race bias in Germany: testing the contact hypothesis with Turks and Germans. Paper presented at the 4th European Conference on Psychology and Law of the American Psychology-Law Association and the European Psychology-Law Association in Dublin, Ireland
- Sporer SL (2001a) Recognizing faces of other ethnic groups: an integration of theories. *Psychol Public Policy Law* 7:36–97
- Sporer SL (2001b) The cross-race bias: beyond recognition of faces in the laboratory. *Psychol Public Policy Law* 7:170–200
- Sporer SL, Cohn LD (2011) Meta-analysis. In: Rosenberg B, Penrod SD (eds) *Research methods in forensic psychology*. Wiley, New York, pp 43–62
- Sporer SL, Horry R (2011) Pictorial versus structural representations of in-group and out-group faces. *J Cogn Psychol* 23:974–984. <https://doi.org/10.1080/20445911.2011.594434>
- Sporer SL, Trinkl B, Guberova E (2007) Matching faces: differences in processing speed of out-group faces by different ethnic groups. *J Cross-Cult Psychol* 38:398–412
- Statistisches Bundesamt (2019) Bevölkerung und Erwerbstätigkeit. Ausländische Bevölkerung. Downloaded Feb 29, 2020, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/auslaend-bevoelkerung-2010200187004.pdf?__blob=publicationFile
- Tredoux C, et al. (2020) 50 years of research on the cross-race effect: a meta-analysis. Paper presented at the American Psychology-Law Conference in New Orleans, LO
- Valentine T (1991) A unified account of the effects of distinctiveness, inversion and race on face recognition. *Q J Exp Psychol* 44:671–703
- Valentine T, Chiroro P, Dixon R (1995) An account of the own-race bias and the contact hypothesis based on a 'face space' model of face recognition. In: Valentine T (ed) *Cognitive and computational aspects of face recognition*. Explorations in face space. Routledge, London, pp 69–94
- Wells GL, Windschitl PD (1999) Stimulus sampling and social psychology experimentation. *Personal Soc Psychol Bull* 25:1115–1125. <https://doi.org/10.1177/01461672992512005>
- Wells GL, Small M, Penrod SJ, Malpass RS, Fulero SM, Brimacombe CAE (1998) Eyewitness identification procedures: recommendations for lineups and photospreads. *Law Hum Behav* 22:603–647
- White D, Kemp RI, Jenkins R, Matheson M, Burton AM (2014a) Passport officers' errors in face matching. *PLoS One* 9:e103510. <https://doi.org/10.1371/journal.pone.0103510>
- Wright DB (2006) Causal and associative hypotheses in psychology: examples from eyewitness testimony research. *Psychol Public Policy Law* 12:190–213
- Wright DB, Sladden B (2003) An own gender bias and the importance of hair in face recognition. *Acta Psychol* 114:101–114
- Wright DB, Stroud JN (2002) Age differences in lineup identification accuracy: people are better with their own age. *Law Hum Behav* 6:641–654
- Wright DB, Boyd CE, Tredoux CG (2003) Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Appl Cogn Psychol* 17:365–373
- Zimmerman LA, Sporer SL (2010) A meta-analytic review of research comparing police to civilian eyewitnesses. Poster presented at the Meeting of the American Psychology-Law Association in Vancouver, Canada