KATEDRA PSYCHOLOGIE

Soutěž psychologických prací
na Katedře psychologie
Fakulty sociálních studií MU
SOUTĚŽNÍ TEXT

# REVERSE-CORRELATING MENTAL REPRESENTATIONS OF ROMANI FACES: THE ROLE OF IMPLICIT PREJUDICE

*Bc. Anna Marie Rosická*

Katedra psychologie, Fakulta sociálních studií, Masarykova univerzita, Brno, Česká republika

anna.rosicka@seznam.cz

**Abstrakt:**

Tato replikace studie Dotsche et al. [Psychological Science, 19, 978–980 (2008)] zkoumala, jak se implicitní předsudky vůči Romům projevují na mentálních reprezentacích romských obličejů. Pomocí metody reverse correlation byly vizualizovány přibližné mentální reprezentace romských obličejů u 34 participantů z řad české majority. Míra implicitních předsudků u těchto participantů byla měřena metodou ST-IAT. U získaných obrázků obličejů byla následně nezávislými participanty ohodnocena míra dvou vlastností běžně spojova-ných s romskými stereotypy – inteligence a kriminálnosti. Vyšší míra implicitních předsudků predikovala signifikantně nižší hodnocení inteligence obličejů, ale pouze marginálně vyšší hodnocení kriminálnosti. Tyto výsledky naznačují, že některé stereotypní vlastnosti se mohou projevovat v mentálních reprezentacích silně-ji než jiné. Nalezené vztahy svým směrem odpovídaly výsledkům replikované studie, ale lišily se v míře prů-kaznosti. Malá velikost nalezených účinků zpochybňuje přesvědčivost závěrů Dotsche et al. (2008) a pouka-zuje na roli různých moderujících faktorů včetně kulturního kontextu. V závěru práce jsou shrnuty limity výzkumu a jeho význam.

**Klíčová slova:** implicitní předsudky, reverse correlation, Romové, vizuální stereotypy, vnímání obličejů

**Abstract:**

This replication of the study by Dotsch et al. [Psychological Science, 19, 978–980 (2008)] examined how im-plicit bias against Romani is manifested in mental representations of Romani faces. Reverse correlation was used to visualise approximate mental representations of Romani faces in 34 Czech majority participants, whose levels of implicit prejudice were measured in a single-target IAT. The obtained classification images were then rated by independent participants on two traits related to the Romani stereotype – intelligence and criminality. Higher levels of implicit prejudice predicted significantly lower intelligence ratings, but only marginally higher criminality ratings of the classification images. These results suggest that some stereotypi-cal attributes can be manifested more strongly in facial mental representations than others. The results were consistent with the replicated study in their direction but not in their conclusiveness. The small size of found effects challenges the robustness of conclusions made by Dotsch et al. (2008) and emphasizes the role of vari-ous moderating factors, including cultural context. Finally, the limitations and implications of the present study are discussed.

**Keywords:** implicit prejudice, reverse correlation, Romani, visual stereotypes, face perception

## INTRODUCTION

In Europe, anti-Romani sentiments are still a pressing issue (EU-MIDIS II, 2016). This study is a contribution to the growing body of literature on visual aspects of ethnic bias. Facial appearance, which affects impression formation and inter-group behaviour (Zebrowitz & Montepare, 2008), is often the first cue for ethnicity judgements. Ethnic bias does not manifest itself only at the explicit verbal level, but also in the form of automatic processing tendencies, i.e. implicitly (Greenwald & Banaji, 1995). For an ecologically valid explanation of intergroup bias and its underlying processes, extending research also to the implicit and visual facets of bias is essential (McArthur & Baron, 1983).

Newly developed image manipulation techniques have recently enabled to visualize the specific facial features linked to various stereotypes. One of these new methods – "reverse correlation" – can help to depict how people envision the faces typical of ethnic groups. Dotsch, Wigboldus, Langner and van Knippenberg (2008) used this method to show that mental representations of ethnic out-group faces can be affected by implicit prejudice against said out-group. In their study, Dotsch and colleagues (2008) visualised how Dutch participants envision a face typical of Moroccans, a highly stigmatized minority in the Netherlands. The resulting pictures suggested that people with higher levels of implicit prejudice have more negatively stereotyped (i.e. more criminal and less trustworthy) mental representations of ethnic out-group faces.

The present study seeks to replicate the findings of Dotsch et al. (2008) in a different ethnic setting, i.e. in the Czech context. The aim here is (a) to introduce reverse correlation as a new way towards understanding stereotypes related to Romani, a widespread yet stigmatized ethnic minority in Czechia; (b) to visualize mental representations of Romani faces; and (c) to examine the link between said representations and implicit prejudice.

## THEORETICAL FRAMEWORK

### Intergroup Bias

To reduce amounts of information from our complex social environment, we tend to categorize others with respect to their group membership (Macrae & Bodenhausen, 2000). Such simplification can occur at the cost of intergroup bias[1]. Along other general information, category membership is extracted from faces more readily than identity-specific information (Quinn &

[1] Although the terms "bias", "prejudice" and "stereotypes" are often used interchangeably, in this study, I am going to differentiate between them similarly to Greenwald et al. (2002). According to this conception, "stereotype" is a cognitive association of attributes with a social group, "prejudice" is a form of attitude (in that it attaches negative valence to the group in question), and the umbrella term "intergroup bias" includes all affective, cognitive and behavioural manifestations of group favouritism.

Macrae, 2011). Category-related knowledge can then be used to derive generalizing judgements which take precedence over individual attributes (Fiske & Neuberg, 1990).

Group membership is sufficient to make us think of others in terms of “us” and “them” and to treat them accordingly (Tajfel, Billig, Bundy & Flament, 1971). We typically differentiate more between the members of our in-group (i.e. the group of people to which we belong) than the members of our out-group (i.e. the group to which we do not belong), which we perceive as more homogenous and stereotypical (Quattrone & Jones, 1980). Moreover, we tend to favour our in-group over the out-group (Dovidio, Hewstone, Glick & Esses, 2010). However, people are not always willing or able to verbalize their stereotypes and attitudes (Greenwald & Banaji, 1995). Implicit measures such as the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) can capture less blatant aspects of attitudes, which would not be accessible through self-report measures (Greenwald & Banaji, 1995).

The unlimited scope of perceptual features possibly involved in social inferences complicates the identification of the visual aspect of stereotypes (Todorov et al., 2011). This can be solved by data-driven techniques such as reverse correlation (Mangini & Biederman, 2004), which, unlike hypothesis-driven approaches, do not rely on the researcher’s judgement as to what features of the stimuli are relevant to social categories (Todorov et al., 2015). Reverse correlation has helped to approximate mental representations of various social groups (Imhoff et al., 2011; Dotsch, Wigboldus & van Knippenberg, 2011; Oldmeadow, Sutherland & Young., 2013), socially relevant traits (Dotsch & Todorov, 2012; Éthier-Majcher, Joubert & Gosselin, 2013) and even particular people (Karremans, Dotsch & Corneille, 2011; Young, Ratner & Fazio, 2014).

Implicit bias about perceived social categories influences further processing of faces (van Knippenberg & Dijksterhuis, 2000). For instance, Hugenberg and Bodenhausen (2004) found that in perceivers with higher levels of implicit prejudice, the readiness to perceive anger was larger for African American than for European American faces, while positive emotion was recognized faster in European American than in African American faces (Hugenberg, 2005). Whether a person is categorized in terms of a category depends on the fit between that person’s characteristics and the characteristics expected from said category (van Knippenberg & Dijksterhuis, 2000). Along these lines, Dotsch and colleagues (2008) argued that people’s expectations about prototypical out-group faces can be affected by their beliefs about what personality traits are typical for the group in question. The study by Dotsch and colleagues (2008) lent support to this hypothesis, in that the estimated mental representations of the faces of racial out-group members proved to be more stereotype-congruent (i.e. more criminal- and less trustworthy-looking) in more prejudiced participants.

### Current Study

While a lot of research on ethnic bias has been conducted in ethnically diverse societies, many regions are much more homogenous with respect to ethnicity of their population, as is the case with Czechia. Out-groups may be perceived differently in a setting in which they are less common. The aim of this study is to see if the findings by Dotsch et al. (2008) apply to the Czech cultural context. This replication is focused at Romani, the most numerous yet very controversial Czech minority. This ethnicity was chosen because its members can be encountered on everyday basis in Czechia, and therefore, its knowledge in the Czech majority is not solely based on media images. Romani children are stereotyped to not comply with the demands of regular formal education, Romani neighbourhoods as dangerous and Romani themselves as criminal (Weinerová, 2014).

In the present study, I expect that implicit prejudice towards Romani will be reflected in the stereotypicality of mental representations of Romani faces. Specifically, the approximate mental representations of participants with higher levels of implicit prejudice against Romani (as measured by a single-target IAT) will be rated by independent raters as less intelligent (H1) and more criminal-looking (H2) than the classification images of less prejudiced participants.

## METHODS

### Participants

For the first part of the study, 34 Czech participants[2] (aged 19 – 27, $M$ = 22.91, $SD$ = 2.17; of that 17 women) were recruited in Brno using convenience and snowball sampling. The sample size was set to closely resemble the sample size in Dotsch et al. (2008, Study 2). Participants were approached in person at the Brno University of Technology and Faculty of Informatics, Masaryk University, or via social media (Facebook study groups, Twitter). Thus, the sample consisted mostly of university students (28/34, i.e. 82%). The admission criteria were age (18 – 27, to prevent age-related slowing in the ST-IAT; Ratcliff, Spieler, & McKoon, 2000) and field of study or work: to achieve more variability in levels of implicit prejudice, humanities students were excluded along with members of people-oriented specializations such as education or law. To further increase the naivety of participants, the study was advertised as 'visualization of mental representations: picturing how we envision the faces of others'.

For the second part of the study, other 104 people were recruited via social media. Of that, 93 participants (aged 18 – 27, $M$ = 22.20, $SD$ = 1.77; of that 63 women) met the admission criteria (similar to Part 1). The groups of participants in each part of the study did not differ significantly in age, $t(52.09) = -1.60$, $p = 0.12$.

### Procedure

Similarly to Dotsch et al. (2008, Study 2), the present study consisted of two parts and used a mixed quasi-experimental design. In the first part, participants were brought to the test site and asked to provide their informed consent (see Appendix 1). This part was administered individually to eliminate possible distractions and social desirability effects. Within the informed consent, participants were told that the study would examine their mental representation of a face typical for a certain group of people, such as a nationality. Next, they filled in short paper-and-pen questionnaires on sociodemographic information (see Appendix 2) and proceeded to the reverse-correlation task, which is described in more detail in the next section. The session ended with a single-target IAT (ST-IAT) which measured implicit bias and which was introduced as "a simple categorization task". After that, the participants were debriefed and thanked.

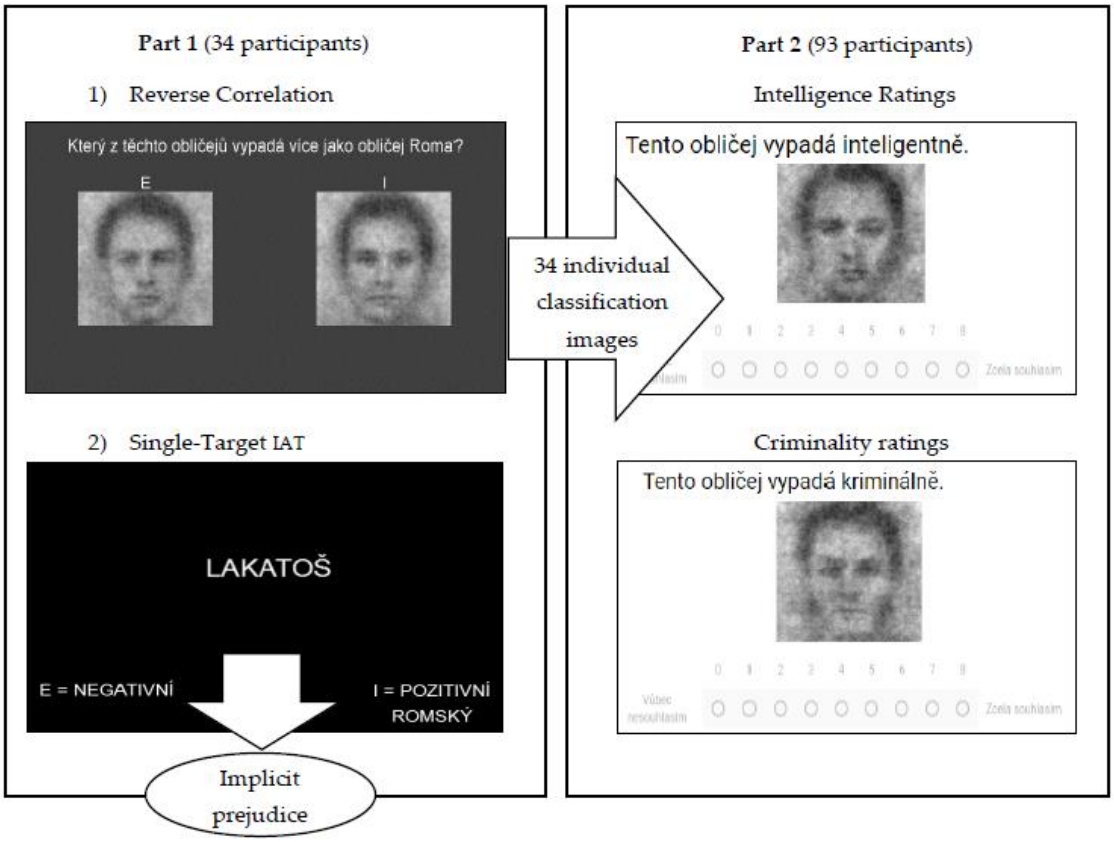
In the second part of the study, the results of the reverse correlation task (34 classification images of faces) were each rated on two stereotype-related traits by 93 independent, hypotheses-

[2] Originally, 36 participants were recruited but two male participants had to be excluded from analyses on the grounds of technical difficulties that disrupted the sessions.

blind participants. Each trait was rated separately in two counterbalanced blocks via an online questionnaire on Google Forms. The image order was randomized within each block to diminish fatigue effects. The traits in question – criminality (a stereotype-congruent trait) and intelligence (a stereotype-incongruent trait) – were rated on a 9-point Likert scale (0 – *strongly disagree* to 8 – *strongly agree*) instead of a 7-point scale as in Dotsch et al. (2008), to ensure enough variability for the use of parametric statistical tests. The procedure is summarised in Fig. 1.

**Figure 1**

*An overview of the procedure.*




## Reverse Correlation

To visualize mental representations of out-group faces, a forced-choice reverse correlation task was used (Mangini & Biederman, 2004; Dotsch et al., 2008). This task was run in PsychoPy (Peirce, 2007), and consisted of 500 trials as compared to the original 770 to keep the participants motivated while still producing valid classification images.

In each trial, participants were presented with two stimuli faces simultaneously and asked to select the face that looked more like a face of a Romani by pressing the keys "E" or "I". After every 50 trials the participants could take a short break. An individual classification image was computed for each participant by averaging all stimuli faces that the participant had chosen as more Romani-like (for examples see Appendix 5).

**Stimuli**

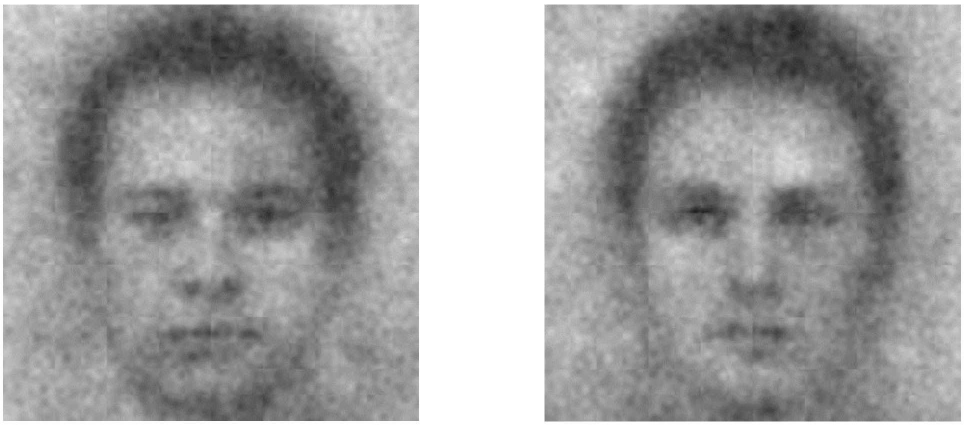
The stimuli faces were obtained using the same base face as Dotsch and colleagues (2008). The 1000 stimuli pictures were generated in R version 3.3.3 (R Core Team, 2017) using the rcicr package by Dotsch (2016), by superimposing random sinusoid noise over the base face (Figs. 1 and 2, for more detail on noise computation see Dotsch et al., 2008). In the task, each stimulus face with a random noise pattern added was paired with a stimulus face in which the same noise pattern was reversed (i.e. subtracted). The set of stimuli pairs was identical for all participants but presented in random order, and the side on which pictures with added or subtracted noise appeared varied randomly as well.

**Figure 2**

*Base face used to generate stimuli for the reverse-correlation task.*

**Figure 3**

*Examples of stimuli faces used in the reverse correlation task.*



**Single-Target Implicit Association Test**

To indirectly measure implicit prejudice, a single target IAT was used (ST-IAT, Bluemke & Friese, 2008; Dotsch & Wigboldus, 2008). The ST-IAT consisted of three blocks, in which words appeared in the centre of the screen and participants were asked to correctly categorize them by pressing the left or right key ("E" or "I"). Each trial was preceded by a 300 ms fixation sign. The stimuli word was then presented and remained onscreen until the participant pressed a key. In case of incorrect responses, the word "ERROR" ("CHYBA" in Czech) was presented for 1000 ms. The participants were instructed to answer as quickly and accurately as possible.

In the first (valence-practice) block that consisted of 20 randomly ordered trials, participants categorized positive and negative words (e.g. "love", "hate" or "cancer") as positive or negative. The stimuli words were translated from the original study by Dotsch and colleagues (2008, as provided by R. Dotsch in personal communication, March 8, 2017), or in untranslatable cases substituted by other words of consideration (for the full list of stimuli words see Appendix 3).

In the other two blocks, one more category ("Romani") was added. Now, participants categorized not only positive and negative words, but also typical Romani surnames (for the full list see Appendix 4). In the stereotype-congruent block, words were categorized as Romani or negative by pressing the left key and as positive by pressing the right key. In the stereotype-incongruent block, this was reversed, in that now words were categorized as Romani or positive by pressing the right key and as negative by pressing the left key. Each block consisted of 40 randomly-ordered trials, in which the number of words per each key (left or right) was equal (10 negative, 10 Romani and 20 positive words in the congruent block or 10 positive, 10 Romani and 20 negative words in the incongruent block). To avoid learning effects, the order of congruent/incongruent blocks was counterbalanced between participants with 16 and 18 participants starting with the congruent and incongruent block respectively.

The response latencies on the congruent vs. incongruent block were compared to indicate the strength of participants' positive or negative associations of Romani surnames. Learning blocks and first trials of each critical block were omitted from the calculations similarly to Bluemke & Friese (2008). Remaining latencies were processed using the scoring algorithm by Greenwald and colleagues (2002), resulting in so-called D-scores. A positive D-score indicates a rather positive association with Romani names; a negative D-score a rather negative association.

**Data Analyses**

The subsequent data analyses were conducted in R using the packages "car" (Fox &Weisberg, 2011) and "lme4" (Bates, Maechler, Bolker & Walker, 2015). To inspect the effects of implicit prejudice on the ratings of classification images, a linear mixed model (LMM) was used, which enabled to control for the variability in ratings across individual raters, as well as for classification-image-specific variability. Thus, apart from examining the impact of D-scores (fixed effects), raters and items (i.e. classification images) were introduced to the model as crossed random effects. All ratings (both on intelligence and criminality) were included into one model by adding the rated trait as a binary interaction term.

The assumptions for the use of a LMM were checked. For the final model, residuals approximated normal distribution (as shown by a P-P plot) and were not substantially autocorrelated (in the Durbin-Watson test, $D = 2.01$). A scatterplot of residuals reflected the nonlinear clustering of the dependent variable. This was understandable given that ratings were measured on a Likert-scale.

## RESULTS

### Descriptive Statistics

The D-scores of participants in the first part of the study were slightly negative on average ($M_{\text{D-score}}$ = - 0.04, $SD$ = 0.374, see Table 1) but not significantly different from zero, $t(33) = 0.69$, $p$ = .495. Neither the Kolmogorov-Smirnov test, nor the Kruskal-Wallis test showed violations of normality. Both statistical and graphical checks (histogram) suggested the distribution of D-scores approximated normal distribution.

Two raters were excluded from analyses because they gave the same rating to all pictures, presumably because they did not try to provide valid responses. From the remaining 91 raters, there were 6188 ratings in total, 3094 for each trait, 2x91 for each classification image. On average, the pictures were rated as rather unintelligent ($M_{\text{intelligence}} = 3.68$) and rather criminal ($M_{\text{criminality}} = 4.05$; see Table 1). Because the pictures were rated on a Likert scale, statistical tests of normality justifiably indicated a highly non-normal distribution. However, the histograms suggested an approximately normal distribution of frequencies of each rating. For the sake of statistical analyses, the ratings were treated as a continuous variable. The average ratings of each picture on both traits were significantly negatively correlated, $r(32) = -.68$, $p < .01$, meaning that the images that were averagely perceived as more intelligent were also perceived as less criminal and vice versa.

**Table 1**

*Summary of the independent variable (D-scores indicating positive or negative implicit prejudice) and the dependent variable (classification image ratings).*

| | *N* | *Mean* | *Median* | *SD* | *Min* | *Max* |
|---|---|---|---|---|---|---|
| **D-score** | 34 | -0.0443 | -0.0957 | 0.374 | −0.976 | 0.655 |
| **Ratings (total)** | 6188 | 3.87 | 4 | 2.02 | 0 | 8 |
| Ratings (criminality) | 3094 | 4.05 | 4 | 2.12 | 0 | 8 |
| Ratings (intelligence) | 3094 | 3.68 | 4 | 1.91 | 0 | 8 |

| **Ratings: Frequencies** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Total | 306 | 521 | 845 | 1045 | 1025 | 1066 | 770 | 355 | 255 |
| Criminality | 155 | 247 | 395 | 470 | 466 | 537 | 422 | 219 | 183 |
| Intelligence | 151 | 274 | 450 | 575 | 559 | 529 | 348 | 136 | 72 |

**Main analyses**

Trellis plots of within-rater linear fits showed that raters differed with respect to trends and variability in their rating tendencies, which justified the use of a LMM with raters as a random effect. Because each rater rated the same set of items, these were added to the model as a second random effect. Since there was only one model for ratings of both intelligence and criminality, the random effects were always assessed in interaction with rated trait. The models were estimated using the maximum likelihood procedure (ML) to enable between-model comparisons of fit. Type II Wald chi-square tests were used to estimate the significance of individual fixed effects.

First, a null model was fit to the data to estimate the proportion of variance in ratings accounted for by individual raters and items (see Table 2, Model 0). Both random effects were included in the model in an interaction with the variable "trait", which enabled to see the influence on criminality and intelligence ratings separately. The residual interclass correlation coefficient (ICC, calculated according to Rabe-Hesketh & Skrondal, 2008), was .33 for raters and .16 for classification images, meaning that within-rater and within-item consistency of ratings explained 33% and 16% of the variance in ratings respectively.

Next, rated trait was added to the model as a fixed effect to take into account the differences in average ratings of each rated trait. Because the order of rating blocks was randomized (41 raters rated intelligence first, 50 criminality first) and histograms had shown more uniform distribution of ratings in the second block, it was desirable to control for order effects. Therefore, the variable "order" was added as another fixed effect, with values of 0 or 1 meaning the rater had rated all items on intelligence or criminality first respectively. The model 1 (see Table 2, Model 1) did not show a significantly better fit compared to the null model, $\chi^2(2) = 3.6207$, $p = .16$, and neither of the fixed effects was significant, $\chi^2_{trait}(1) = 2.19$, $p = 0.14$; $\chi^2_{order}(1) = 1.60$, $p = .21$. However, the variables "order" and "trait" were kept in the model as control variables.

Finally, a full model was fit (see Table 2, Model 2) by adding a "trait:D-score" interaction to model 1 as the main fixed effect. This enabled to look at effects of implicit prejudice specific to criminality and intelligence ratings. Adding this predictor did not significantly increase the fit of the model, $\chi^2 (2) = 5.6463$, $p = .06$. The effects of trait and order remained nonsignificant, $\chi^2_{trait}(1) = 2.41$, $p = .12$; $\chi^2_{order}(1) = 1.59$, $p = .21$, but the interaction of trait and D-score was significant on overall, $\chi^2_{trait:D\text{-}score}(2) = 6.14$, $p < .05$. Thus, an increase in D-score by 1 predicted non-significantly lower ratings on criminality ($b = -0.60$, $SE = 0.33$, 95% CI [-1.25; 0.05]) and significantly higher ratings on intelligence ($b = 0.59$, $SE = 0.24$, 95% CI [0.12; 1.06]). These results support the hypothesis that higher levels of implicit prejudice predict lower intelligence ratings of one's classification images. However, the hypothesis that higher levels of implicit prejudice predict higher criminality ratings of one's classification image was not supported, although the effect was directed in the predicted direction. Each of the three models explained approximately 34% of data variance ($\Omega^0{}_2$ = .34, calculated according to Xu, 2003). Thus, although the predicted effects were present, they were marginal in size.

**Table 2**

*Model specifications and fixed effects estimates (top). Variance-covariance estimates (bottom).*

| | Model 0 | | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| Formula: | Rating ~ 1 + | | | Rating ~ 1 + trait + | | | Rating ~ 1 + trait + | | |
| | + (0 + trait \| rater) + | | | + order + | | | + order + trait:Dscore + | | |
| | + (0 + trait \| item) | | | + (0 + trait \| rater) + | | | + (0 + trait \| rater) + | | |
| | | | | + (0 + trait \| item) | | | + (0 + trait \| item) | | |
| **Fixed Effects** | B | SE | CI | B | SE | CI | B | SE | CI |
| Intercept | 3.82*** | 0.08 | [3.67; 3.98] | 3.95*** | 0.19 | [3.59; 4.3 | 3.93*** | 0.18 | [3.57; 4.28] |
| Trait | | | | -0.37 | 0.25 | [-0.85; 0.12] | -0.31 | 0.24 | [-0.78; 0.15 |
| Order | | | | 0.17 | 0.14 | [-0.10; 0.44] | 0.17 | 0.14 | [-0.10; 0.44 |
| Trait:Dscore | | | | | | | | | |
| *For criminality ratings* | | | | | | | -0.60 | 0.33 | [-1.25; 0.05 |
| *For intelligence ratings* | | | | | | | 0.59* | 0.24 | [0.12; 1.06] |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 2.79 (1.67) | | | 2.79 (1.67) | | | 2,79 (1.67) | | |
| $\tau^2_{rater}$ – *criminality ratin* | 1.19 (1.09) | | | 1.14 (1.07) | | | 1.14 (1.07) | | |
| $\tau^2_{rater}$ – *intelligence ratings* | 0.56 (0.75) | | | 0.57 (0.75) | | | 0.57 (0.75) | | |
| $\tau^2_{item}$ – *criminality ratin* | 0.55 (0.74) | | | 0.52 (0.72) | | | 0,47 (0.69) | | |
| $\tau^2_{item}$ – *intelligence ratings* | 0.30 (0.54) | | | 0.28 (0.53) | | | 0.23 (0.48) | | |
| $ICC_{rater}$ | 0.32 | | | | | | | | |
| $ICC_{item}$ | 0.16 | | | | | | | | |
| | | | | | | | | | |
| -2LL | -12255.90 | | | -12252.48 | | | -12252.48 | | |
| AIC | 24527.81 | | | 24524.96 | | | 24524.96 | | |
| BIC | 24581.65 | | | 24524.96 | | | 24592.26 | | |
| $\chi 2$ | | | | 3.62 | | | 5.65 | | |

*Note*: * $p < 0,05$; ** $p < 0.01$; *** $p < 0.001$. The SDs for random effects are in parentheses. For the dummy variable “trait”, 0 = intelligence and 1 = criminality; for the dummy variable “order”, 0 = intelligence first, 1 = criminality first.

## DISCUSSION

The aim of this study was to investigate how interethnic prejudice manifests itself in visual mental representations of ethnic out-group and to replicate the findings by Dotsch and colleagues (2008) in another cultural context. Reverse correlation was used to visualize approximate mental representations of Romani faces in a Czech sample. The classification images of more prejudiced participants were expected to look more criminal and less intelligent than the classification images of less prejudiced participants.

The results of this study suggest that higher levels of implicit prejudice towards Romani predict significantly less intelligent, but only marginally more criminal mental representation of Romani faces. However, the size of both effects is very small. The direction of found effects is consistent with the replicated study, but there is a discrepancy in strength of these effects. The original study by Dotsch et al. (2008) found that implicit prejudice predicted trustworthiness and criminality ratings of the mental representations. In the present study however, only one of the stereotype-related traits was linked to implicit prejudice.

The failure to fully replicate the findings by Dotsch et al. (2008) could hint at a real intercultural difference. It is possible that the levels of bias against Romani in Czechs are smaller than the levels of bias against Moroccans in the Dutch. Moroccan and Romani stereotypes are also likely to differ in their content. What is central to the Morrocan stereotype may be only trivial to the Romani stereotype and vice versa. Romani faces could also differ from Czech faces less than Moroccan faces do from the Dutch, and so the group-related stereotype could be encoded more strongly in other, non-visual attributes.

The discrepancy may also arise from different statistical approaches (using a LMM rather than computing many linear regressions). Participants in the first part of the study could differ in how they approached the reverse correlation task: some participants may have systematically chosen faces that were less deformed by noise, because regardless of their ethnicity, these looked more like real faces. Consequently, the resulting classification images may differ in other respects than the amount of bias they reflect. Dotsch and colleagues did not systematize the part of error explained by item characteristics. This could have led to an overestimation of the effect size. The present study, then, might offer a more accurate depiction of reality in this respect.

## Limitations and Future Directions

The main limitation may lie in a low number of observations. Although the number of raters in the present study was almost twice as large as in the original study by Dotsch et al. (2008), the number of within-subject trials on the reverse correlation task may have been insufficient to get valid approximations of mental representations. There were only 500 trials as compared to the 770 trials in the original study, which may have caused inaccurate classification images. The resulting classification images were all highly similar to each other (for a few examples, see Appendix 5, Fig. 4). Thus, the raters may have been less able to discern slight traces of bias in the rated images.

Another problem may lie in the use of Romani surnames as target stimuli on the ST-IAT. It is possible that not all participants were familiar with presented surnames, and therefore the stimuli did not always fully activate the target attitude. In such case, participants could classify any novel-sounding or unusual words as Romani without paying attention to the target category. This could have decreased the validity of D-scores.

# CONCLUSION

The current findings lent partial support to the conception that implicit interethnic bias manifests itself in mental representations of faces. The results suggest that higher levels of implicit bias against Romani are linked to mental representations of Romani faces as less intelligent, but insufficient evidence was found to conclude that said bias also predicts more criminal mental representations of Romani faces. Although the found effect was very small, the present findings still make an important contribution to the literature on visual stereotype, in that the data challenge the robustness of conclusions made by Dotsch et al. (2008). Moreover, these findings indicate that the content of visual stereotypes about ethnic out-group differs across ethnic contexts.

**REFERENCES:**


Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, *38*(6), 977-997.

Dotsch, R. (2016). *rcicr: Reverse correlation image classification toolbox*. R package version 0.3.4.1.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3 (5), 562-571

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19*, 978–980.

Dotsch, R., & Wigboldus, D. H. (2008). Virtual prejudice. *Journal of experimental social psychology*, *44*(4), 1194-1198.

Dotsch, R., Wigboldus, D. H., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of personality and social psychology*, *100*(6), 999.

Dotsch, R., Wigboldus, D. H., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology, 43*, 116–125

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: theoretical and empirical overview. In J. F. Dovidio, V. M. Esses, P.Glick, & M. Hewstone (Eds.), *The SAGE handbook of prejudice, stereotyping and discrimination* (pp. 3-29). London: SAGE Publications Ltd.

Éthier-Majcher, C., Joubert, S., & Gosselin, F. (2013). Reverse correlating trustworthy faces in young and older adults. *Frontiers in psychology*, *4*, 592.

*EU-MIDIS II: Second European Union Minorities and Discrimination Survey. Roma–Selected findings* [Report]. (November 2016). Retrieved from http://fra.europa.eu/en/publication/2016/eumidis-ii-roma-selected-findings

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in experimental social psychology*, 23, 1-74.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage. Retrieved from: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review, 102*(1), 4-27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, *109*(1), 3.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A., Poehlman, T., Uhlmann, E., & Banaji, M. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. J*ournal Of Personality And Social Psychology*, 97(1), 17-41.

Hugenberg, K. (2005). Social categorization and the perception of facial affect: target race moderates the response latency advantage for happy faces. *Emotion, 5*, 267–276.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: the role of prejudice and facial affect in race categorization. *Psychological Science, 15*, 342–345.

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous ingroup projection. *Psychological Science*, *22*, 1583-1590.

*Inquisit 5* [Computer software]. (2016). Retrieved from http://www.millisecond.com

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, *121*(3), 422-426.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology*, *51*(1), 93-120.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, *28* (Rendering the Use of Visual Information from Spiking Neurons to Recognition), 209-226.

Oldmeadow, J. A., Sutherland, C. A., & Young, A. W. (2013). Facial stereotype visualization through image averaging. *Social Psychological and Personality Science*, *4*(5), 615-623.

Payne, B. K., Jacoby, L. L., & Lambert, A. J. (2005). Attitudes as accessibility bias: Dissociating automatic and controlled processes. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.). *The new unconscious* (pp. 393-420). Oxford; New York : Oxford University Press.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.

Quattrone, G. A., Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology, 38*, 141–152.

Quinn, K. A., & Macrae, C. N. (2011). The face and person perception: Insights from social cognition. *British Journal of Psychology*, *102*(4), 849-867.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. Texas: STATA press.

Ratcliff, R., Spieler, D., McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin and Review*, 7, 1-25.

Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the implicit association test. *Group Processes & Intergroup Relations*, 10(3), 359-372.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal Of Social Psychology*, *1*(2), 149.

Todorov, A., Dotsch, R., Wigboldus, D. J., & Said, C. P. (2011). Data-driven Methods for Modeling Social Perception. *Social & Personality Psychology Compass*, *5*(10), 775-791.

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current opinion in neurobiology*, *23*(3), 373-380.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.

van Knippenberg, A., & Dijksterhuis, A. (2000). Social categorization and stereotyping: A functional perspective. *European Review of Social Psychology*, *11*(1), 105–144.

Weinerová, R. (2014). *Romové a stereotypy*. Prague: Karolinum Press.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of experimental social psychology*, *10*(2), 109-120.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in medicine*, *22*(22), 3527-3541.

Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias the mental representation of a presidential candidate’s face. *Psychological Science*, *25*(2), 503-510.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497-1517.

Appendix 1

**Projekt: Vizualizace mentálních reprezentací**

Fakulta sociálních studií

Joštova 10, 602 00 Brno

Masarykova univerzita v Brně

Kontaktní osoba: Anna Marie Rosická

Tel: 605 430 014, E-mail: anna.rosicka@seznam.cz

**Informovaný souhlas**

Vážená paní, vážený pane,

děkuji, že jste si na mě udělal(a) čas! Jsem studentkou psychologie na brněnské Masarykově univerzitě a ve své bakalářské práci se zabývám mentálními reprezentacemi obličejů, tedy tím, jak si představujeme tváře druhých. Obracím se na Vás s prosbou o účast na výzkumu, ve kterém se pokusím pomocí jednoduchého počítačového úkolu vizualizovat, jak si představujete tvář typickou pro určitou skupinu lidí, např. národnost.

Výzkum je rozdělen na tři části. V té první Vás poprosím o vyplnění některých základních demografických údajů. Následuje hlavní část výzkumu, vizualizace mentálních reprezentací, která je poměrně časově náročná, abychom získali co nejvýstižnější obrázek Vaší mentální reprezentace. Poslední částí je jednoduchý a krátký úkol na třídění slov. Celkem výzkum zabere přibližně půl hodiny.

Všechny informace o sobě, které mi v rámci výzkumu poskytnete, zůstanou anonymní. Vaše data budou použita výhradně pro účel výzkumu. Kdykoli v průběhu máte možnost odmítnout pokračovat a z experimentu odstoupit. V takovém případě budou Vaše informace skartovány.

Pokud máte nějaké otázky, neváhejte se prosím zeptat.

Pokud souhlasíte s účastí na tomto výzkumu, přečtěte si a podepište prosím následující prohlášení.

Prohlašuji, že

souhlasím s účastí ve výzkumu Anny Marie Rosické. Jsem obeznámen(a) s průběhem studie a souhlasím, aby všechny získané údaje o mé osobě byly použity pro výzkumné účely. Jsem si vědom(a), že výsledky výzkumu mohou být anonymně publikovány. Jsem informován(a), že mám možnost kdykoliv od spolupráce na projektu odstoupit, a to i bez udání důvodu.

Jméno, příjmení a podpis účastníka v projektu:

_______________________________________________________________________________

________________________________V_____________________dne:_________________

Appendix 2

**Participant č. ____**

**Dotazník**

1. Jsem:
    a. Muž
    b. Žena
2. Věk:____
3. Nejvyšší dosažené vzdělání:

_______________________________________________________

4. Jsem:
    a. Student – uveďte prosím obor: ____________________
    b. Pracující – uveďte prosím své zaměstnání: ____________________
    c. Jiné – doplňte: ____________________

Appendix 3

**Table 3**

*Positive and negative stimuli words used in the ST-IAT. Both Czech originals and English translations are listed.*

| **Positive words** | | **Negative words** | |
|---|---|---|---|
| Czech | English | Czech | English |
| LÁSKA | Love | RAKOVINA | Cancer |
| MÍR | Peace | VÁLKA | War |
| BEZPEČÍ | Safety | NEŠTĚSTÍ | Disaster |
| ZDRAVÍ | Health | SMRT | Death |
| RADOST | Joy | MUČENÍ | Torture |
| MILÝ | Kind | KATASTROFA | Catastrophe |
| VESELÝ | Cheerful | SMUTEK | Sadness |
| ŠTĚSTÍ | Happiness | NEHODA | Accident |
| POTĚŠENÍ | Delight | ZTRÁTA | Loss |
| ÚSPĚCH | Success | BOLEST | Pain |
| PŘÍJEMNÝ | Pleasant | HNUS | Disgust, filth |
| VYHRÁT | Win | NENÁVIST | Hate |
| ZÁBAVA | Fun | ZLO | Evil |

| | | | |
|---|---|---|---|
| PŘÍTEL | Friend | NEMOC | Illness |
| SMÍCH | Laughter | OTRAVNÝ | Annoying |
| ÚCTA | Respect | OŠKLIVÝ | Ugly |
| PĚKNÝ | Nice | NEBEZPEČÍ | Danger |
| HEZKÝ | Pretty | ODPORNÝ | Disgusting, nasty |
| BLAHO | Bliss | TYRAN | Tyrant |
| DOBRO | Good, welfare | UTRPENÍ | Suffering |

Appendix 4

**Table 4**

*List of Romani surnames used in the ST-IAT task.*

| |
|---|
| OLÁH |
| HORVÁTH |
| NEMÉTH |
| MIRGA |
| DŽUGI |
| BADI |
| LAKATOŠ |
| ŽIGO |
| BALOG |
| BADŽO |

Appendix 5

**Figure 4**

*Examples of classification images – approximated mental representations of Romani faces.*