Social Categorization Based on Language and Facial Recognition

Marie-France Champoux-Larsson¹, Frida Ramström¹, Albert Costa²,³, and Cristina Baus⁴

Abstract
Two experimental studies were conducted to replicate the effect found by Baus et al. where language as a marker of social categories affected recognition of faces in an old/new paradigm. In Study 1, we presented faces along with utterances in Swedish and in English to native Swedish speakers. Faces presented along with Swedish utterances were not recognized better than faces presented along with English utterances. In Study 2, we used another language pair and presented faces along with utterances in Swedish and in Spanish to native Swedish speakers. Faces presented along with Swedish utterances were recognized better than faces presented along with Spanish utterances. Our results suggest that language functions as a marker of social categories and that, similarly to other markers of social categories, it can be modulated by various factors and is not unconditional.

Keywords
social categorization, language, recognition, memory, ingroup/outgroup

Social categorization is one of the most investigated phenomenon in social psychology (Hugenberg & Sacco, 2008). Social categorization is a process through which the amount of information that the brain processes is reduced by clustering stimuli into groups based on a common characteristic (e.g., Ellemers & Haslam, 2012; Sheepers &

¹Mid Sweden University, Östersund, Sweden
²Universitat Pompeu Fabra, Barcelona, Catalunya, Spain
³Institució Catalana de Recerca i Estudis Avancats, Barcelona, Catalunya, Spain
⁴University of Barcelona, Catalunya, Spain

Corresponding Author:
Marie-France Champoux-Larsson, Department of Psychology and Social Work, Mid Sweden University, Campus Östersund, 83125 Östersund, Sweden.
Email: mfclarsson@gmail.com
Ellemers, 2019; Tajfel, 1982). Since categorizing individuals into social groups recruits fewer cognitive resources than perceiving individuals as discrete entities does, more of the brain’s limited resources remain available and can be assigned to other cognitive tasks (Macrae et al., 1994). Although a person belongs to several social categories simultaneously, for instance, based on their gender, ethnicity, age, or even profession or favorite sports team, we tend to categorize an individual based on the most relevant or salient category within a given context (e.g., Macrae et al., 1995; Stroessner, 1996), and/or based on our personal motivations and preconceptions (e.g., Fein & Spencer, 1997; Hugenberg & Bodenhausen, 2004). Indeed, categorizing an individual based on all possible characteristics simultaneously would defeat social categorization’s purpose of saving cognitive resources (Hugenberg & Sacco, 2008).

As we categorize people into social categories, we tend to amplify the similarities that members of outgroups share, and to magnify between-group differences (Hugenberg & Sacco, 2008; Tajfel & Wilkes, 1963). As a consequence of this, less attention is paid to outgroup members’ uniqueness (Ellemers & Haslam, 2012), and we perceive ingroup members with more precision. When observing outgroup members, common and stereotypical characteristics shared by all (or most) members are sought, as opposed to searching for distinctive features when observing ingroup members (Ellemers & Haslam, 2012; Hugenberg & Sacco, 2008). This difference in how we perceive and encode outgroup and ingroup members has effects on other cognitive processes, such as memory and recognition. Several studies show that we more accurately recognize ingroup members’ faces than outgroup members’ faces. This effect has been found for gender (for a meta-analysis, see Herlitz & Lovén, 2013), ethnicity (for a meta-analysis, see Meissner & Brigham, 2001), age (for a meta-analysis, see Rhodes & Anastasi, 2012), and even characteristics that are not related to physical attributes such as college affiliation or alleged personality type (e.g., Bernstein et al., 2007).

Although social categorization can occur based on virtually any characteristic (or markers), gender, ethnicity, and age are among the most important social categorization markers that we use and they have been investigated extensively (for meta-analyses of each respective marker, see Herlitz & Lovén, 2013; Meissner & Brigham, 2001; Rhodes & Anastasi, 2012). Recently, language use (including accents) has been increasingly investigated in relation to social categorization, and empirical findings suggest that language is also used as a marker of social categories by both children (e.g., Byers-Heinlein et al., 2017; Kinzler & Dautel, 2012; Kinzler & DeJesus, 2013; Kinzler et al., 2007, 2012) and adults (e.g., Hansen et al., 2014, 2016; Lev-Ari & Keysar, 2010; Pietraszewski & Schwartz, 2014). Interestingly, language patterns may be even more important than other major markers (such as ethnicity) in social interactions. For instance, Kinzler et al. (2009) found that when 5-year-old children choose to be friends with other children, they prefer different-ethnicity children with a native accent to same-ethnicity children without a native accent. In adults, Rakić et al. (2011) found similar results where participants categorized people as being native or foreigner based on their accent rather than based on their ethnicity when they had access
to both cues simultaneously. More recently, Paladino and Mazzurega (2020) also showed that accent is more important than ethnicity when adult form concepts of ingroup and outgroup. These results suggest that language is a major marker of social categorization that is more important than other social markers during social interactions.

Importantly, because we encode outgroup members in a more stereotypical manner, another consequence of social categorization is that it can lead to stereotyping, which in turn can lead to prejudice and even discrimination (Brown, 2000; Cuhadar & Dayton, 2011; Schaller & Neuberg, 2008). Since attitudes toward a specific language and speakers of that same language correlate strongly (for an overview, see Dragojevic et al., 2021), hearing a foreign language is likely to give rise to associated stereotypes toward the speaker as well. In the globalized world that we live in today, contact with people speaking a foreign language (or with a foreign accent) is probably more common than ever before. The possibility that prejudice and discrimination can ensue from social categorization based on language is therefore an imminent risk. Indeed, there is already evidence for discrimination and negative attitudes toward foreign speakers and foreign-accented speakers (e.g., Birney et al., 2020; Hansen et al., 2017; Hendriks et al., 2018; Roessel et al., 2018, 2019, 2020). Understanding social categorization based on language and its underlying mechanisms is thus imperative in order to prevent and fight potential ramifications such as prejudice and discrimination (Schaller & Neuberg, 2008).

Nevertheless, the effect that social categorization based on language has on face recognition has not received much attention yet. A notable exception is a study by Baus et al. (2017) where they found that faces that were presented along with utterances in the participants’ native language (Spanish) were recognized more accurately than faces that were presented along with utterances in the participants’ foreign language (English). To our knowledge, this was the first study to demonstrate the effects of social categorization on face recognition when using language as a social marker. Yet, as suggested by the Self-Categorization theory, categories are variable and highly dependent on what is relevant for the observer in a given time and place (Turner & Reynolds, 2012). Although the effects of social categorization on face recognition are generally universal and robust, they are also known to vary based on a variety of factors such as culture and context (for meta-analyses, see Bettencourt et al., 2001; Fischer & Derham, 2016). Indeed, what constitute a relevant social group in one cultural context may not necessarily do so in another culture. Consequently, although the English speakers in Baus et al. (2017) could be perceived as an outgroup, it cannot be automatically assumed that they would be perceived as such in another culture, or that language will lead to an ingroup bias on memory in a different context. Furthermore, in the light of the widely discussed replication crisis in psychological science, replication studies with appropriate statistical power are recommended (e.g., Anderson & Maxwell, 2017; Maxwell et al., 2015; Shrout & Rodgers, 2018), perhaps even more so for newly found effects that are not robustly established. Therefore, this proposed effect should be tested more thoroughly in other populations and using different language pairs.
The aim of the present study was thus to expand the results in Baus et al. (2017) by investigating whether the effect would replicate in individuals with the same foreign language (i.e., English), but with a different native language (i.e., Swedish) in Study 1. In Study 2, we investigated whether the effect would be replicated when Spanish was used as the foreign language in a population of native Swedish speakers. In both studies, we used an old/new paradigm. The old/new paradigm is used extensively to test other types of own-group biases (e.g., Brown et al., 2017; Herzmann et al., 2017; Michel et al., 2006; Stelter & Degner, 2018), including in the study by Baus et al. (2017). A typical old/new task (Sporer, 2001) begins with an encoding phase during which faces from the participant’s ingroup and from their outgroup are shown. During this phase, participants are asked to pay attention to the faces, but are not informed that they will have to recall the faces later on. After a distraction task (incorporated in order to avoid any recency effect), a surprise recognition phase follows, during which both old (i.e., presented earlier during the encoding phase) and new (i.e., presented for the first time) faces are presented. During the surprise recognition phase, participants are asked to determine whether the face that they see was presented in the first part of the experiment (i.e., encoding phase), or whether it is presented for the first time. The ingroup bias is found in terms of a larger number of faces recognized for ingroup faces than for outgroup faces (for a meta-analysis, see Herlitz & Lovén, 2013). The methods and results along with a short discussion for Study 1 and Study 2 respectively are presented together before a general discussion where the results of both studies are discussed together.

**Study 1**

**Method**

**Participants.** A total of 65 participants ($M_{age}=27.7, SD=8.7; 65\%$ females) took part in this part of the study. However, three participants were excluded since they reported not having Swedish as a first language, three were excluded because they reported having English as a first language, five for reporting having lived abroad in an English speaking country for more than 6 months, and one participant was excluded for reporting not having normal or corrected hearing. The final sample thus included 53 participants ($M_{age}=28.1, SD=9.4; 62\%$ females). A power analysis using the effect size ($\eta_p^2=.038$) from Baus et al. (2017) showed that 43 participants were sufficient to have an 80\% chance of finding a similar effect size. Participants were recruited amongst students on the university’s campus via ads and personal contacts. The most frequent completed education level was high school (Mode=2, $n=48$: on a scale where 1=elementary school or lower, 2=high school, 3=professional education, 4=Bachelor’s degree, 5=Mastér’s degree, 6=PhD). Participants reported their level of proficiency in Swedish and English respectively by answering five questions (in general, written understanding, oral understanding, written production, oral production) on a scale from 1 to 10 where lower scores represent lower proficiency. An average of the five questions was computed to create a language proficiency score. The average language
skill for Swedish was 9.4 (SD=0.8) and 6.9 (SD=1.5) for English. The difference between the Swedish and English scores was significant, \( t(52)=13.71, p < .001, d=1.88 \).

**Materials.** A set of 80 pictures were selected from the Chicago Face Database (Ma et al., 2015). Pictures of neutral facial expressions with the highest ratings on prototypicality and with the lowest ratings on unusualness were selected (based on norming data provided in Ma et al., 2015). To remove distracting features such as hair and skin redness, pictures were cropped into an oval to include the face from hairline to chin, and from ear to ear. Pictures were also turned into shades of black and gray. All pictures depicted adult Caucasians where half were males and half were females.

The same 40 sentences that were used in Baus et al. (2017) were used in this study. The sentences were simple every day utterances such as “Those shoes are new,” “That door was open,” and “The sky is cloudy.” For the Swedish stimuli, the sentences were translated from English to Swedish and recorded using 20 different native Swedish speakers (10 of each gender). The English sentences were also recorded using 20 native English speakers from North America (American or Canadian accent, 10 of each gender). All 40 sentences were recorded from all 20 Swedish and all 20 English speakers. Sentences were uttered in a neutral tone of voice and recorded with a Zoom H2n Handy Recorder microphone. The recordings were prepared in Audacity by removing silences before and after the utterances and by normalizing the frequency across recordings. The decibel level of all files was then normalized using Levelator.

The sentences in Swedish and English did not differ in terms of number of words \( (M_{\text{Swedish}}=4.68, SD_{\text{Swedish}}=1.05; M_{\text{English}}=4.93, SD_{\text{English}}=1.05), t < 1 \), but the length of the utterances (in ms) differed \( (M_{\text{Swedish}}=2,000, SD_{\text{Swedish}}=442; M_{\text{English}}=1,647, SD_{\text{English}}=361), t(158)=5.54, p < .001, d=0.88 \) (see the interim discussion below about this issue).

Participants also filled out a questionnaire with demographic questions (gender, age, highest completed education level, whether the participant had lived in an English speaking country and if so for how long). Furthermore, the questionnaire contained a question as to whether the participant had a normal (or corrected) vision and hearing. Proficiency questions for Swedish and English (as described under Participants), as well as the lexical test LexTALE (Lemhöfer & Broersma, 2012) were also presented through this questionnaire. Note however that LexTALE was included in the questionnaire for the purposes of a student paper and that the score was not used in this study.

**Design and procedure.** Information about the study was provided and written informed consent was collected. Afterwards, the participants completed the experiment. Instructions for the experiment were all presented in writing. The experiment consisted of two phases: an encoding phase and a recognition phase. During the encoding phase, pictures of faces were presented along with a recorded utterance in Swedish or English. Trials were presented for 2,700 ms to accommodate the longest recording, and were followed by a cross that appeared in the center of the screen for 500 ms. Presentation was semi-randomized into four different versions to which participants were
assigned randomly. Across the four different versions, gender, language, and sentences were counterbalanced and presented evenly so that each sentence was uttered by a male native speaker, a female native speaker, a male foreign speaker, and a female foreign speaker across the versions, and each sentence was presented only once within the same version of the experiment. Furthermore, pictures were evenly paired with native and foreign speakers across the versions. Pictures that were presented in the encoding phase and recall phase were also counterbalanced across the versions. Within each version of the experiments, a voice, a sentence, and a picture were presented only once, but all voices, sentences, and pictures were presented in all versions. The voice’s and face’s gender were always congruent. During the encoding phase, 20 male and 20 female faces and voices were presented, of which half were in Swedish, and half were in English (evenly distributed across gender). The instructions given to the participants for the encoding phase were to pay close attention to the faces and utterances.

Between the encoding and recognition phases, a distraction task that lasted for approximately 7 minutes was given to the participants to reduce the risk of a recency effect. A mental rotation task using letters presented normally or mirrored was used. A trial started with a fixation cross in the center of the screen for 500 ms after which a letter appeared on the screen. The participant’s task was to determine whether the letter was mirrored or not, and the next trial started after the participant provided an answer, or after 3,000 ms if no answer was given. A total of 70 trials were presented.

A surprise recognition phase followed during which the 40 pictures presented during the encoding phase along with 40 new pictures (half were females) were presented randomly. Participants were asked to answer whether they remembered the face from the first part of the experiment, or whether it was a new face. There was no time limit to provide an answer, and each trial was separated by a fixation cross presented for 500 ms. Afterwards, the sentences used in the encoding phase along with 40 new sentences (half in Swedish) were presented in writing. Here again, the participants had to determine whether the sentence was old (i.e., heard in the first phase), or new. Each trial was separated by a fixation cross presented for 500 ms. The experiment was programmed and presented with E-Prime 2.0 (Psychology Software Tools, 2012).

When the participant was finished with the experimental part, they filled out the survey online. They received information about the full purpose of the study, given a movie ticket to thank them for their participation, along with written information about the study and contact information. All national and international ethics laws and recommendations were followed.

Data preparation and analyses. For accuracy, signal detection theory methods (Stanislaw & Todorov, 1999) were used to code the answers from the recognition phase as was done in Baus et al. (2017). Namely, answers were coded into hits (old stimuli correctly identified as old), misses (old stimuli falsely identified as new), correct rejections (new stimuli correctly identified as new), and false alarms (new stimuli falsely identified as old). A sensitivity score, $d'$, was computed for each condition (Swedish, English) based on the z-transformed ratio of hits minus the z-transformed ratio of false alarms ($d' = z[\text{hits}] - z[\text{false alarms}]$). Accuracy ($d^*$) was tested using both a classical
paired sample one-tailed $t$-test and a Bayesian paired sample one-tailed $t$-test (including inferential plots of prior and posterior distribution, robustness check, and sequential analysis, all reported in the Supplemental Materials). The prior for the Bayesian analysis was set to the program’s standard value of 0.707 due to the scarcity of knowledge on the effect, and the interpretation of the Bayesian factor was based on the guidelines in Wagenmakers et al. (2018). Furthermore, a sensitivity score was also computed for the sentences in Swedish and the sentences in English in the recognition phase using the same procedure. The same analyses with the same settings as for the faces were used. All analyses were conducted in JASP (JASP Team, 2020). Note that the analyses that we conducted differed slightly from Baus et al. in two ways. First, we conducted $t$-tests on accuracy for the native and foreign language conditions rather than conducting an ANOVA for accuracy for the native language condition, the foreign language condition, as well as correct rejections. Given that the correct rejections are part of the accuracy measurement in terms of false alarms, testing the difference between the native and foreign language conditions was deemed appropriate for the purpose of this study. Also, another difference is that we conducted both classical and Bayesian analyses on our data in order to draw more robust conclusions about our results.

**Results and Discussion**

The one-tailed $t$-test for accuracy for the pictures showed that there was no difference on $d'$ between the Swedish ($M=0.399$, $SD=0.066$) and English faces ($M=0.397$, $SD=0.072$) in the recognition phase, $t(52)=0.223$, $p=0.412$, $d=0.031$ (see the left panel in Figure 1). The Bayesian one-tailed $t$-test ($BF_{10}=0.18$) confirmed that there was no evidence for an effect in our data. The robustness of the Bayes factor was good, with most prior widths leading to moderate or strong evidence for the null hypothesis (see Figures 1SM–3SM in the Supplemental Materials).

As for the sentences, data from three participants was excluded since their hit rate and/or their false alarm rate was either 1 or 0 in at least one condition (which hinders the computation of $d'$). The one-tailed $t$-test was significant indicating that Swedish sentences ($M=0.44$, $SD=0.09$) were better recognized than English sentences ($M=-0.37$, $SD=0.19$), $t(49)=23.74$, $p<0.001$, $d=3.36$ (see the left panel in Figure 2). As for the Bayesian one-tailed $t$-test, it provided strong evidence that participants recognized the Swedish sentences better than the English sentences as well ($BF_{10}>100$). The robustness of the Bayes factor was high, with all except the narrowest prior widths leading to decisive evidence in favor of the alternative hypothesis (see Figures 3SM and 4SM in the Supplemental Materials).

Language ingroup faces were not recognized better than language outgroup faces in Study 1. This result is inconsistent with Baus et al. (2017) and with other studies where social categorization occurred based on language use (e.g., Hansen et al., 2014, 2016; Lev-Ari & Keysar, 2010; Pietraszewski & Schwartz, 2014). Interestingly, Swedish and English as a language pair has previously yielded inconsistent results with other types of paradigms. Such an example of divergent results is found with the
Foreign language effect, where one tends to make more rational or utilitarian decisions when using a foreign language than when using a native language (for an overview, see Costa et al., 2017, and for a meta-analysis, see Circi et al., 2021). For instance, while the Foreign language effect has been found and replicated in several studies using several language pairs, no effect was found when Dylman and Champoux-Larsson (2020) attempted to replicate the effect in a Swedish population using English as a foreign language. In contrast, the effect occurred when they used Swedish and French as a language pair. Dylman and Champoux-Larsson suggested that the lack of effects when using Swedish and English could be due, at least in part, to the way Swedes learn English. More specifically, Swedes tend to learn English in a meaningful sociocultural context rather than in a sterile class environment (for a discussion on how the sociocultural learning environment of a foreign language affects emotional connections when using it, see Caldwell-Harris, 2014; Dewaele, 2008, 2015). The authors proposed that the mechanisms that usually lead to a Foreign language effect were not activated when English was used in a Swedish population since English probably has stronger emotional connections and cannot be considered to be a “true” foreign language for Swedes.
Nonetheless, the analysis of the sentences clearly shows that participants remembered Swedish sentences significantly better than English sentences. This suggests that foreign utterances were not processed as effortlessly as native language utterances. Furthermore, participants showed a strong false alert rate for the foreign sentences as shown by the negative $d'$ average for English utterances. This indicates that our participants either paid less attention to foreign utterances, or had more difficulty encoding them.

A limitation in our design was that the length of the sentences was significantly longer in Swedish than they were in English. Although faces were presented for the same duration regardless of the language they were presented along with, we cannot exclude the possibility that participants spent more time encoding the faces presented along with English than the faces presented along with Swedish utterances since they were not busy listening to English utterances for as long as for Swedish utterances. If so, this could explain why we did not find any difference in recognition between the Swedish and English conditions. Yet, participants’ recognition of Swedish faces was not impeded by the fact that they simultaneously paid attention to and encoded Swedish utterances better.

**Figure 2.** Mean sensitivity scores ($d'$) for the texts in the two studies. 
*Note.* The panel on the left shows the mean sensitivity scores ($d'$) for the texts in the first study in the Swedish and English conditions. The panel on the right shows the mean sensitivity scores ($d'$) for the texts in the second study for the Swedish and Spanish conditions.
Nevertheless, given the special status that English has in Sweden, we conducted a second study with the same design, but with a language pair that did not pose the same concerns as English. Specifically, we used Spanish as a foreign language in a population of native Swedish speakers. Spanish as a foreign language is not as common as English in Sweden and does not have the same cultural importance (European Commission, 2012). Furthermore, for those who speak Spanish as a foreign language in Sweden, their skills tend to be lower than their English skills (Skolverket, 2012). Therefore, the relationship between Swedish and Spanish in Sweden is more comparable to the relationship between Spanish and English in Spain.

Study 2

Method

Participants. A total of 29 participants ($M_{\text{age}} = 32.6, SD = 13.5; 59\%$ females) took part in the second part of the study. However, one participant was excluded since they reported not having Swedish as a first language, and another one for reporting Spanish as a first language. The final sample thus included 27 participants ($M_{\text{age}} = 33.2, SD = 13.8; 63\%$ females). Although this was lower than the sample size of 43 participants that would have been optimal for an $80\%$ chance of finding the same effect size as in Baus et al. (2017), data collection had to be terminated to follow the social restrictions directed by Swedish authorities due to the outbreak of the Covid-19 pandemic. Furthermore, our final sample size was similar, albeit slightly smaller, to the sample size in Baus et al. (i.e., $N = 33$). Participants were recruited via ads on the department’s social media and on the university’s electronic bulletin board, as well as through announcements in forums for speakers of Spanish as a foreign language in Sweden. The most frequent education level (based on the same scale as in Study 1) was Master’s degree (Mode = 4, $n = 14$). Participants reported their level of proficiency in Swedish and Spanish respectively by using the same scales as in Study 1 and the language proficiency scores were computed in the same manner. The average language skill for Swedish was $9.5 (SD = 0.6)$ and $3.5 (SD = 2)$ for Spanish. The difference between the Swedish and Spanish scores was significant, $t(26) = 16.27, p < .001, d = 3.13$.

Material. The materials used were identical to the ones used in Study 1 with the exception of the sentences in English, which were replaced with sentences in Spanish. The same recordings as those used in Baus et al. (2017) were presented in this part of the study. The sentences in Swedish and Spanish did not differ in terms of number of words ($M_{\text{Swedish}} = 4.68, SD_{\text{Swedish}} = 1.05; M_{\text{Spanish}} = 4.73, SD_{\text{Spanish}} = 0.75$), $t < 1$. To overcome the limitation of Study 1 where we discovered a significant difference in length between the utterances in the native and foreign languages, we modified the length of the utterances in Swedish to make them equivalent to the Spanish utterances. The Swedish recordings were slightly accelerated, but without changing the pitch or making them sound unnatural. The length (in ms) of the utterances in both
languages was thus equivalent ($\text{MSwedish} = 1,485$, $\text{SD}_{\text{Swedish}} = 312$; $\text{MSpanish} = 1,484$, $\text{SD}_{\text{Spanish}} = 272$), $t < 1$.

**Design and procedure.** The design and procedure were identical to the ones used in Study 1, except for the stimuli that were in English in Study 1 were in Spanish in this study.

**Data preparation and analyses.** Data were prepared using the same steps as in Study 1.

**Results and Discussion**

The one-tailed $t$-test showed that there was a difference on $d'$ between the Swedish ($M = 0.43$, $SD = 0.11$) and Spanish faces ($M = 0.34$, $SD = 0.09$) in the recognition phase, $t(26) = 4.11$, $p < .001$, $d = 0.79$ (see the right panel in Figure 1). The Bayesian one-tailed $t$-test confirmed this difference as well, suggesting that our data provided strong evidence for the effect of language ($BF_{10} = 172.39$). The robustness of the Bayes factor was high, with all except the narrowest prior widths leading to decisive evidence in favor of the alternative hypothesis (see Figures 6SM–8SM in the Supplemental Materials).

As for the sentences, data from four participants was excluded since their hit rate and/or their false alarm rate was either 1 or 0 in at least one condition. The subsample for the analyses for the sentences thus consisted of 23 participants ($M_{\text{age}} = 34.2$, $SD = 14.7$; 61% females) with the most frequent education level being Master’s degree (Mode = 4, $n = 12$). The subgroup’s average language skill for Swedish was 9.6 ($SD = 0.5$) and 3.3 ($SD = 2.2$) for Spanish. The one-tailed $t$-test revealed that sentences in Swedish ($M = 0.55$, $SD = 0.11$) were recognized better than sentences in Spanish ($M = -0.37$, $SD = 0.13$), $t(22) = 24.5$, $p < .001$, $d = 5.11$ (see the right panel in Figure 2). The Bayesian one-tailed $t$-test confirmed this with strong evidence as well ($BF_{10} > 100$). The robustness of the Bayes factor was high, with all except the narrowest prior widths leading to decisive evidence in favor of the alternative hypothesis (see Figures 9SM and 10SM in the Supplemental Materials).

When using Swedish and Spanish as a language pair, the effect of social categorization based on language on recognition was replicated from Baus et al. (2017). Namely, participants were more likely to remember a face when it had been paired with a Swedish utterance in the encoding phase than when it had been paired with a Spanish utterance. Although our sample was relatively small, the Bayesian analysis, which is not dependent on sample sizes in the same way as classical analyses are (see Schönbrodt et al., 2017), provided very strong evidence for an effect of language on memory.

Interestingly, as in the Swedish-English sample, Swedish sentences were remembered better than the sentences in the foreign language, and a strong false alarm rate was found for the Spanish utterances. In other words, the results from Study 1 and Study 2 for the sentences were comparable and therefore suggest that foreign language sentences were processed in a similar way regardless of the foreign language. Specifically, both foreign languages were more difficult to process than Swedish. It also suggests that the sentences were processed in the same way across the studies.
despite the duration difference between Swedish and the foreign language in Study 1. However, whether the lower performance for foreign sentences represents reduced attention or more difficulty understanding or encoding the sentences cannot be determined with this design.

**General Discussion**

In this study, we investigated whether we could replicate the findings in Baus et al. (2017), where social categorization based on language led to an ingroup bias in face recognition, using different language pairs. In Study 1, we found that when English was used as a foreign language, as in Baus et al., but in a sample of Swedish native speakers, no ingroup bias occurred. However, the effect from Baus et al. was replicated when Spanish was used as a foreign language in Study 2, also in a group of Swedish native speakers. Importantly, in both studies, participants recognized utterances in their native language much better than the utterances in the foreign language, suggesting that both English and Spanish were processed as foreign languages. Interestingly, even though the sentences were processed in a similar way in both Study 1 and in Study 2, the effect of social categorization on recognition of faces was significant only in Study 2. Although more research will be needed to determine the cause of this difference between the language pairs, we suggest a few tentative explanations that may be worth exploring in the future.

A first possible explanation is the difference in proficiency in the foreign languages. The average proficiency in English ($M=6.9$) appears to be much higher than the average proficiency in Spanish ($M=3.5$). Yet, the hit rate pattern for the foreign sentences compared to the native sentences was virtually identical across the studies, suggesting that proficiency in and of itself is not sufficient to explain the different results between Study 1 and Study 2. If proficiency on its own would have been the reason for the different results, we would have expected the participants to remember English utterances as well as the Swedish ones, which they did not. Nevertheless, future research should explore the importance of proficiency further in order to establish whether and how it may affect social categorization. For instance, a study where multilingual participants are tested in their native language, in a high proficient foreign language, and in a low proficient foreign language could provide important insights on this issue.

Another explanation may lay in the difference in experience that our participants have with each respective language rather than to proficiency per se. Admittedly, exposure to and proficiency in a foreign language tend to go hand in hand (e.g., De Cat, 2020; De Wilde et al., 2019; Place & Hoff, 2011; Unsworth et al., 2015, 2018). However, because of the results found for the sentences, a possibility is that the amount of contact with the foreign language led to the different results in the two samples. Indeed, research suggests that contact of good quality with an outgroup (which Swedes might have with English) can reduce or even eliminate ingroup biases (Hancock & Rhodes, 2008; Rhodes et al., 2009; Sporer, 2001; Wan et al., 2015). In Sweden, English has a high status and a prominent place in society (Kulturdepartementet, 2008), as well as a significant cultural prevalence through several media (television, movies, music, ads,
literature, etc.) coming directly from English-speaking countries and being consumed in their original form (Kulturdepartementet, 2002). Thus, although English is a foreign language for most Swedes, many have meaningful regular contacts with the language. It is therefore possible that the Swedes’ regular contacts with the English language created the conditions necessary to reduce the outgroup bias in face recognition.

Additionally, an important aspect that should be investigated further is the fact that our participants in Study 2 most likely spoke Spanish as a third language, given that English is taught in Swedish school as a compulsory subject early on during education. Therefore, we cannot completely exclude the possibility that the first foreign language will not lead to social categorization, but that subsequent foreign languages will. Indeed, participants in Baus et al. (2017) spoke English as one of three (or more) languages, with Spanish and Catalan being the two other ones. On the other hand, both Spanish and Catalan were likely native languages for the participants.

As for how social categorization based on language can lead to stereotyping, prejudice, and discrimination, future research should address several facets of this issue. For example, there is a need to investigate specifically which attitudes and prejudice are associated with various languages across different cultures. Indeed, research shows that attitudes toward a specific language group vary depending on the immigration pattern in a given area or country (e.g., Asbrock et al., 2014). Taking the socio-cultural and political context into consideration will therefore be important. Another fundamental mission for future research will be to determine how to prevent negative stereotypes, prejudice, and discrimination from arising as a result of perceiving individual as an outgroup based on their language. It will be important to determine in which specific circumstances the ingroup bias emerges and what the underlying mechanisms are. By understanding the underpinnings of social perception and categorization based on language, interventions to prevent and eliminate potential ramifications can be developed and targeted to appropriate groups, thus reducing such negative consequences (Dragojevic et al., 2021).

To sum up, our results suggest that language is a marker of social categories, but that, as other markers of social categories, it is not unconditional and unchangeable. This goes in line with other research suggesting that language is an important marker of social categories, but that it can be modulated with experience with outgroup members (for meta-analyses, see Pettigrew & Tropp, 2006, 2008). On a theoretical point of view, this study suggests that language should receive more attention as a marker of social categories. On the practical plan, given that social categorization can lead to stereotyping, which in turn can lead to prejudice and discrimination (Allport, 1979; Harris & Fiske, 2006; Liberman et al., 2017), it is important to understand how to prevent negative consequences that social categorization based on language could potentially lead to, especially given the international context that we live in today.

**Acknowledgments**

The authors would like to thank Stefan Becker and Andrew Karlsson for their contribution with the creation of stimuli and for their assistance with data collection, and the two anonymous reviewers for their helpful feedback on an earlier version of this article.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Cristina Baus was supported by the Ramon y Cajal research program (RYC2018-026174-I).

ORCID iD
Marie-France Champoux-Larsson https://orcid.org/0000-0001-7251-5263

Supplemental Material
Supplemental material for this article is available online.

References
Allport, G. W. (1979). The nature of prejudice. Basic Books.
Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “replication crisis”: Using original studies to design replication studies with appropriate statistical power. Multivariate Behavioral Research, 52(3), 305–324. http://dx.doi.org/10.1080/00273171.2017.1289361
Asbrock, F., Lemmer, G., Becker, J. C., Koller, J., & Wagner, U. (2014). “Who are these foreigners anyway?” The content of the term foreigner and its impact on prejudice. SAGE Open, 4(2), 1–8. https://doi.org/10.1177/2158244014532819
Baus, C., Bas, J., Calabria, M., & Costa, A. (2017). Speak my language and I will remember your face better: An ERP study. Frontiers in Psychology, 8, 709. https://doi.org/10.3389/fpsyg.2017.00709
Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. Psychological Science, 18(8), 706–712. https://doi.org/10.1111/j.1467-9280.2007.01964.x
Bettencourt, B. A., Charlton, K., Dorr, N., & Hume, D. L. (2001). Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. Psychological Bulletin, 127(4), 520–542. https://doi.org/10.1037//0033-2909.127.4.520
Birney, M. E., Rabinovich, A., Morton, T. A., Heath, H., & Ashcroft, S. (2020). When speaking English is not enough: The consequences of language-based stigma for nonnative speakers. Journal of Language and Social Psychology, 39(1), 67–86. https://doi.org/10.1177/0261919419890430
Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. European Journal of Social Psychology, 30(6), 745–778. https://doi.org/10.1002/1099-0992(200011/12)30:6<745::AID-Ejsp24>3.0.co;2-o
Brown, T. I., Uncapher, M. R., Chow, T. E., Eberhardt, J. L., & Wagner, A. D. (2017). Cognitive control, attention, and the other race effect in memory. PLoS One, 12(3), e0173579. https://doi.org/10.1371/journal.pone.0173579
Byers-Heinlein, K., Behrend, D., Said, L., Girgis, H., & Poulin-Dubois, D. (2017). Monolingual and bilingual children’s social preferences for monolingual and bilingual speakers. Developmental Science, 20(4), e12392. https://doi.org/10.1111/desc.12392
Caldwell-Harris, C. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology, 5*, 1055. https://doi.org/10.3389/fpsyg.2014.01055

Cirri, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*. Advance online publication. https://doi.org/10.3758/s13423-020-01871-z

Costa, A., Vives, M.-L., & Corey, J. D. (2017). On language processing shaping decision making. *Current Directions in Psychological Science, 26*(2), 146–151. https://doi.org/10.1177/0963721416680263

Cuhadar, E., & Dayton, B. (2011). The social psychology of identity and inter-group conflict: From theory to practice. *International Studies Perspectives, 12*(3), 273–293. https://doi.org/10.1111/j.1528-3585.2011.00433.x

De Cat, C. (2020). Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition, 42*(2), 279–325. https://doi.org/10.1017/S0272263119000597

De Wilde, V, Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition, 23*(1), 171–185. https://doi.org/10.1017/S1366728918001062

Dewaele, J.-M. (2008). The emotional weight of I love you in multilinguals’ languages. *Journal of Pragmatics, 40*(10), 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dewaele, J.-M. (2015). From obscure echo to language of the heart: Multilinguals’ language choices for (emotional) inner speech. *Journal of Pragmatics, 87*, 1–17. https://doi.org/10.1016/j.pragma.2015.06.014

Dragojevic, M., Fasoli, F., Cramer, J., & Rakić, T. (2021). Toward a century of language attitude research: Looking back and moving forward. *Journal of Language and Social Psychology, 40*(1), 60–79. https://doi.org/10.1080/0261927X20966714

Dylman, A. S., & Champoux-Larsson, M.-F. (2020). It’s (not) all Greek to me: Boundaries of the foreign language effect. *Cognition, 196*, 104148. https://doi.org/10.1016/j.cognition.2019.104148

Ellemers, N., & Haslam, A. (2012). Social identity theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 379–398). Sage.

European Commission. (2012). *Europeans and their languages* (Special Eurobarometer 386). https://data.europa.eu/data/datasets/s1049_77_1_ebs386?locale=en

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*(1), 31–44. https://doi.org/10.1037/0022-3514.73.1.31

Fischer, R., & Derham, C. (2016). Is in-group bias culture-dependent? A meta-analysis across 18 societies. *SpringerPlus, 5*(70), 1–9. https://doi.org/10.1186/s40064-015-1663-6

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99*(1), 45–56. https://doi.org/10.1348/000712607X199981

Hansen, K., Rakić, T., & Steffens, M. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology, 33*(1), 68–77. https://doi.org/10.1177/0261927X13499761

Hansen, K., Rakić, T., & Steffens, M. (2017). Competent and warm? How mismatching appearance and accent influence first impressions. *Experimental Psychology, 64*(1), 27–36. https://doi.org/10.1027/1618-3169/a000348
Hansen, K., Steffens, M., Rakić, T., & Wiese, H. (2016). When appearance does not match accent: Neural correlates of ethnicity-related expectancy violations. *Social Cognitive and Affective Neuroscience, 12*(3), 507–515. https://doi.org/10.1093/scan/nsw148

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme outgroups. *Psychological Science, 17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Hendriks, B., van Meurs, F., & Reimer, A.-K. (2018). The evaluation of lecturers’ non-native-accented English: Dutch and German students’ evaluations of different degrees of Dutch-accented and German-accented English of lecturers in higher education. *Journal of English for Academic Purposes, 34*, 28–45. https://doi.org/10.1016/j.jeap.2018.03.001

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition, 21*(9–10), 1306–1336. https://doi.org/10.1080/135506285.2013.823140

Herzmann, G., Minor, G., & Adkins, M. (2017). Neural correlates of memory encoding and recognition for own-race and other-race faces in an associative-memory task. *Brain Research, 1655*, 194–203. https://doi.org/10.1016/j.brainres.2016.10.028

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science, 15*(5), 342–345. https://doi.org/10.1111/j.0956-7976.2004.00680.x

Hugenberg, K., & Sacco, D. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass, 2*(2), 1052–1072. https://doi.org/10.1111/j.1751-9004.2008.00090.x

JASP Team. (2020). JASP (Version 0.14) [Computer software]. https://jasp-stats.org/

Kinzler, K. D., & Dautel, J. (2012). Children’s essentialist reasoning about language and race. *Developmental Science, 15*(1), 131–138. https://doi.org/10.1111/j.14677687.2011.01101.x

Kinzler, K. D., & DeJesus, J. M. (2013). Children’s sociolinguistic evaluations of nice foreigners and mean Americans. *Developmental Psychology, 49*(4), 655–644. http://dx.doi.org/10.1037/a0028740

Kinzler, K. D., Dupoux, E., & Spelke, E. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). ‘Native’ objects and collaborators: Infants’ object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development, 13*(1), 67–81. https://doi.org/10.1080/15248372.2011.567200

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children’s social preferences. *Social Cognition, 27*(4), 623–634. https://doi.org/10.1521/ soco.2009.27.4.623

Kulturredarommet. (2002). Mål i mun – Förslag till handlingsprogram för svenska språket [Languages in the mouth – Proposal for an action program for the Swedish language]. https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2002/04/sou-200227/

Kulturredarommet. (2008). Värna språken – förslag till språklag [Protecting languages – proposal for a language law]. https://www.regeringen.se/49bb98/contentassets/80e573241b8e4cc5955510810065f59e/varna-spraken—forslag-till-spraklag-sou-200826

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0
Lev-Ari, S., & Keysar, B. (2010). Why don’t we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The origins of social categorization. *Trends in Cognitive Sciences, 21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Ma, D., Correll, S., & Wittenbrink, J. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macrae, C., Bodenhausen, G., & Milne, A. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*(3), 397–407. https://doi.org/10.1037/0022-3514.69.3.397

Macrae, C., Milne, A., & Bodenhausen, G. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*(1), 37–47. https://doi.org/10.1037/0022-3514.66.1.37

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist, 70*(6), 487–498. https://doi.org/10.1037/a0039400

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037//1076-8971.7.1.3

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14*(1), 55–73. https://doi.org/10.1080/13506280500158761

Paladino, M. P., & Mazzurega, M. (2020). One of us: On the role of accent and race in real-time in-group categorization. *Journal of Language and Social Psychology, 39*(1), 22–39. https://doi.org/10.1177/0261927X19884090

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922–934. https://doi.org/10.1002/ ejsp.504

Pietraszewski, D., & Schwartz, A. (2014). Evidence that accent is a dedicated dimension of social categorization, not a byproduct of coalitional categorization. *Evolution and Human Behavior, 35*(1), 51–57. http://dx.doi.org/10.1016/j.evolhumbehav.2013.09.005

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence two-year-olds’ bilingual proficiency. *Child Development, 82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Psychology Software Tools. (2012). E-Prime [Version 2.0]. https://www.pstnet.com

Rakić, T., Steffens, M., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology, 100*(1), 16–29. https://doi.org/10.1037/a0021522

Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology, 100*(4), 717–728. https://doi.org/10.1348/000712608X396503

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146–174. https://doi.org/10.1037/a0025750
Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*, 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition, 144*, 91–115. https://doi.org/10.1016/j.cognition.2015.07.011

**Author Biographies**

**Marie-France Champoux-Larsson** received her PhD in Psychology at Mid Sweden University, in 2018, where she investigated the perception of emotions in speech and faces in bilingual children and adults. Her current research focuses on how we perceive, categorize, and judge foreign speakers and foreign-accented speakers, how we make decisions in a foreign language, and how we perceive emotions when interacting in foreign language.

**Frida Ramström** received her Masters of Science in Psychology at Mid Sweden University in 2020. She is now working as an Intern Psychologist as school psychologist in the Stockholm area. Her work focuses among other things on social issues regarding immigration, segregation, and subcultures in relation to social categorization based on language in a high school environment.

**Albert Costa** was an ICREA research professor and director of the Speech Production and Bilingualism Research Group at the Centre for Brain and Cognition at Pompeu Fabra University, Barcelona. His research encompassed several areas related to the cognitive and neurological basis of bilingualism, including language processing, attentional control, decision-making, and aging.

**Cristina Baus** is a Ramon y Cajal researcher at the Department of Cognition, Development and Educational Psychology at the University of Barcelona. Her research interests focus on understanding bilingual language processing and its impact beyond the linguistic domain.