Bilingualism: Language and Cognition

cambridge.org/bil

How first- and second-language emotion words influence emotion perception in Swedish–English bilinguals

Marie-France Champoux-Larsson[1,2] and Erik C. Nook[3]

[1]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden; [2]Department of Psychology, University of Chicago, Chicago, USA and [3]Department of Psychology, Princeton University, Princeton, USA

Research Article

Cite this article: Champoux-Larsson, M.-F., & Nook, E.C. (2024). How first- and second-language emotion words influence emotion perception in Swedish–English bilinguals. *Bilingualism: Language and Cognition*, **27**, 939–949
https://doi.org/10.1017/S1366728923000998

Received: 4 August 2023
Revised: 29 November 2023
Accepted: 21 December 2023
First published online: 19 January 2024

Keywords:
second language; emotion perception; bilingualism; facial expressions of emotion; emotion words

Corresponding author:
Marie-France Champoux-Larsson;
E-mail: mfclarsson@gmail.com

Abstract

Emotional experiences are often dulled in one's second language. We tested whether emotion concepts are more strongly associated with first language (L1) than second language (L2) emotion words. Participants (140 L1-Swedish–L2-English bilinguals) saw a facial expression of an emotion (cue) followed by a target, which could either be another facial expression, an L1 emotion word, or an L2 emotion word. Participants indicated whether the cue and target represented the same or different emotions as fast as possible. Participants were faster and more accurate in both the L1 and L2 word conditions compared to the face condition. However, no significant differences emerged between the L1 and L2 word conditions, suggesting that emotion concepts are not more strongly associated with L1 than L2 emotion words. These results replicate prior research showing that L1 emotion words speed facial emotion perception and provide initial evidence that words (not only first language words) shape emotion perception.

Introduction

People often report that their emotions feel dulled when constrained to their second language rather than their first language (see Dewaele, 2010; Pavlenko, 2005, for seminal accounts on the subject). Research on this phenomenon (referred to in the bilingualism literature as reduced emotional resonance [e.g., Harris et al., 2003; Weimer et al., 2022]) has repeatedly shown decreased emotional reactivity when reading, speaking, or listening to material in a second language. For example, when asked to rate positive and negative words on a scale from 'not positive at all' to 'extremely positive' and from 'not negative at all' to 'extremely negative' respectively, bilinguals rate both emotion-label words (e.g., joy, anger) and emotion-laden words (e.g., death, darling) as less emotionally intense in their second (L2) compared to their first (L1) language (e.g., Ferré et al., 2022, for an overview, see Pavlenko, 2008). Similarly, bilinguals judge swear words (Dewaele, 2004), taboo words (Dewaele, 2004; Harris et al., 2003), childhood reprimands (Harris et al., 2003), and saying 'I love you' (Dewaele, 2008) as less emotionally arousing when in L2 than L1. At the behavioral level, reading comprehension tasks show that bilinguals take significantly longer to read L1 negative emotion words than L1 neutral words, but this difference is not observed when comparing L2 negative and neutral words (Sheikh & Titone, 2016). Physiologically, people exhibit lower skin conductance responses (e.g., Caldwell-Harris & Ayçiçeği-Dinn, 2009; Harris, 2004; Harris et al., 2003), smaller pupillary responses (Toivo & Scheepers, 2019), and weaker grip force responses (Thoma et al., 2023) when hearing or reading L2 emotional words compared to L1. Thus, across self-report, behavioral, and physiological levels of analysis, data converge to suggest that emotional experiences are dulled when using one's second language (however, see e.g., Ayçiçegi & Harris, 2004; Ayçiçegi-Dinn & Caldwell-Harris, 2009; Ferré et al., 2010 for studies that did not find such a difference on memory tasks and Eilola & Havelka, 2011; T. M. Sutton et al., 2007 for studies that did not find such a difference on an emotional Stroop task).

Interestingly, converging evidence suggests that L2-proficiency moderates the extent to which emotional resonance is dampened in a second language (e.g., Caldwell-Harris et al., 2011; Harris, 2004). Additional evidence for this notion comes from studies using affective priming tasks, where participants see two emotion words in a row and must determine whether the second word (target) is positive or negative while ignoring the first word (prime) that was presented. Reaction times are faster when the prime is congruent with the target for L1 words but only for L2 words in bilinguals who use their L2 frequently and/or who have high L2-proficiency (e.g., Degner et al., 2012; Winskel, 2013). For less proficient

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

bilinguals, the priming effect weakens or disappears entirely when the task is completed in their less proficient L2. Similarly, it normally takes longer to name the color of negative emotion words than it takes to name the color of neutral words in an (L1) emotional Stroop task. For early bilinguals (i.e., those who acquired both languages during early childhood) that are equally proficient in both their languages (e.g., Grabovac & Pléh, 2014; Sutton et al., 2007) and/or for immersed bilinguals (e.g., Eilola & Havelka, 2011; Grabovac & Pléh, 2014), the emotional Stroop effect still occurs when the task is completed in their L2. However, the effect disappears in participants who have lower L2-proficiency (i.e., they can name the color of negative L2 emotion words just as quickly as the color of L2 neutral words). Consequently, it appears that a dampening of L2 emotional resonance is especially pronounced for individuals with lower L2-proficiency.

Theoretically, constructionist accounts of emotion (Barrett, 2006, 2016) might explain reduced emotion resonance in second language contexts as due to an underlying difference in how closely EMOTION CONCEPTS are associated with L1 vs. L2 emotion words. In brief, constructionist theories posit that emotion words and their associated emotion concepts (i.e., the physiological states, contextual information, and semantic knowledge that together form a person's internal representation of what defines a given emotion) play a central role in shaping both one's OWN emotional experiences and perceptions of OTHERS' emotions (e.g., Barrett, 2006, 2016; Barrett et al., 2007; Gendron et al., 2012; Lindquist & Gendron, 2013; Lindquist et al., 2015). For example, learning to identify novel facial movements as discrete emotions in 'alien' faces (Doyle & Lindquist, 2018) or in chimpanzee faces (Fugate et al., 2010) is facilitated when the novel emotion is paired with an emotion label. In addition, priming the concept of 'fear' (rather than 'anger') before inducing negative affect leads people to behave more fearfully (Lindquist & Barrett, 2008), and having people categorize their emotions using discrete labels shifts neural activity to align with those labels (Satpute et al., 2016). Conversely, when emotion words are unavailable (either due to semantic aphasia or by 'satiating' the meaning of emotion words), others' emotion expressions are perceived mainly in terms of valence (i.e., positive or negative), not discrete emotions (e.g., Doyle et al., 2021; Lindquist et al., 2006, 2014). Consequently, if emotion words (and their underlying emotion concepts) are used when people construct their emotional experiences and perceptions of others' emotions, it is possible that differences in emotional resonance in L2 contexts are due to a reduced association between L2 emotion words and underlying emotion concepts. In particular, bilinguals' semantic network may require them to 'translate' L2 emotion words into L1 emotion words before accessing the emotion concept associated with that L1 emotion word. This theory would also explain improved emotional resonance in high L2-proficiency individuals because this 'translation' may occur more rapidly or it may not be needed at all. Indeed, higher L2-proficiency may overlap with larger experience of the L2, thus increasing the probability that L2 emotion words have been used in an emotionally relevant context. This would lead these individuals to have stronger associations between L2 emotion words and underlying emotion concepts.

Although the research summarized above overall shows reduced emotional resonance in L2 (especially when L2-proficiency is low), we are lacking direct illustrations that this emerges due to weakened activation of emotion concepts by L2 words. Fortunately, a prior study by Nook et al. (2015) developed a paradigm for testing the strength of the association between emotion concepts and L1 emotion words within the context of emotion perception. In two studies, L1 English-speaking participants completed trials in which they first saw a facial expression of an emotion, followed by a target that was either another facial expression (face-face trials) or an English emotion word (face-word trials). Cues and targets were either congruent (i.e., the same emotion, like sad and sad) or incongruent (i.e., different emotions, like sad and angry). Participants indicated whether these pairs expressed the same or different emotions as quickly and accurately as possible within a very short time frame (i.e., 1 second). Both experiments in this study showed that participants' responses were faster and more accurate when targets were emotion words than when targets were other faces. Furthermore, congruent face-word pairs more strongly sped emotion categorization than congruent face-face pairs. These results support the notion that emotion concepts are more strongly ASSOCIATED with emotion words than faces.

Here, we adapted this paradigm to investigate the role of L1 and L2 emotion words on emotion perception. In particular, we added a third condition such that participants completed both congruent and incongruent trials with THREE target types: (i) face-face, (ii) face-L1 word, and (iii) face-L2 word. Comparing participants' accuracy and reaction time across conditions can reveal how strongly people associate conceptual categories of emotion with each of the target types (i.e., faces, L1 words, and L2 words). This is because participants must bring to mind the concept for the cue (e.g., sad) and then indicate whether or not that matches the target category (e.g., sad or angry). Based on the logic of conceptual priming (Collins & Loftus, 1975; Neely, 1991), we can compare reaction times for congruent and incongruent trials across each condition to test how strongly each target type is associated with corresponding emotion concepts. For example, if viewing a sad facial expression speeds the processing of the L1 Swedish emotion word ('sorg') more than the L2 English emotion word ('sad'), that would provide evidence that the concept for sadness is more strongly associated with the L1 than the L2 emotion word. This strengthened association would also produce more accurate categorizations under time pressure.

We used this design to test three questions. First, we tested whether the effects found in Nook et al. replicated in a native Swedish-speaking population. As was found in Nook et al. (2015), we hypothesized that we would again find that the association between congruent cue-target pairs would be stronger (i.e., responses to the target would be faster and more accurate) for face-L1 word pairs than face-face pairs. Second, we tested whether this association would be diminished for emotion words presented in English, our participants' second language. To test whether a weaker association with L2 emotion words may be a mechanism underlying accounts of the reduced L2 emotional resonance described above, we hypothesized that emotions would be perceived more slowly and less accurately in the second language context compared to the first language context and that congruent face-L1 word pairs would result in faster responses than congruent face-L2 word pairs. Finally, we tested whether the participants' L2-proficiency level would moderate these effects. Based on the empirical findings described above suggesting that higher L2-proficiency correlates with higher L2 emotional resonance, we hypothesized that the more proficient participants were in their second language (English), the more facilitating the L2 word context would be for emotion perception. Finding that associations with emotion concepts differ between L1 vs. L2 emotion words could provide a theoretical explanation for the effects reviewed

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

above. Furthermore, addressing this question has important implications for a variety of social situations where interactions occur in a second language, including immigrants living in a country where they are beginning to learn and use a second language, or in contexts where the use of a second language is common such as in business, education, traveling, and online interactions.

Method

Participants

Data for this study were collected in-person in Sweden. Participants were recruited via ads on university bulletin boards and social media. The ad specified which characteristics the participant should have (i.e., Swedish as a L1, English as a L2, age between 18 and 45 years), but no screening was conducted prior to data collection. A total of 154 participants completed the study in-person between late 2019 and early 2022. Based on the same preregistered exclusion criteria presented above, three participants were excluded for having English as one of their first languages (or reported starting to learn English before age 4, see, e.g., Heredia & Cieślicka, 2014; Kovelman et al., 2008), three were excluded for being older than 45 years (see Der & Deary, 2009 for the effect of age on reaction times and Ruffman et al., 2008 for the effect of age on accuracy), and eight were excluded for having mean reaction times or sensitivity scores that were more than 3 standard deviations from the sample mean for one or more conditions (in accordance with our preregistration). Finally, we excluded one participant who failed to respond to 50% of trials in the trial's time limits (average proportion of missing trials for the remaining sample was 0.94%). Although this was not a preregistered exclusion, doing so did not change any results but nonetheless seems prudent for including only high-quality data.

The final sample size was consequently 140 participants (45 [32.14%] male, 93 [66.43%] female, and 2 [1.43%] other; age ranged 18-44 years, $M_{age} = 28.75$ years, $SD_{age} = 5.94$ years, racial demographics were not assessed due to IRB restrictions on collecting these data in Sweden). Education level was distributed with peaks at both high school and bachelor's levels (1 [0.71%] elementary school or lower, 65 [46.43%] high school, 7 [5%] professional training, 46 [32.86%] bachelor's degree, 17 [12.14%] master's degree, 4 [2.86%] Ph.D.). Participants reported their proficiency in Swedish (L1) and English (L2) by rating their skill level on five scales (general proficiency, written understanding, oral understanding, written production, oral production) from 1 to 10 where larger scores indicate higher skill levels. The mean of the five scales was calculated for each participant and language. A paired sample t-test (one-tailed) confirmed that participants' self-reported skills in their L1 ($M = 9.60$, $SD = 0.59$) were significantly higher than their self-reported skills in their L2 ($M = 7.84$, $SD = 1.31$), $t(139) = 17.44$, $p < .001$, $d = 1.59$. As would be expected, self-reported proficiency in L1 and L2 correlated positively, $r(138) = .41$, $p < .001$, 90% CI = [.26, .54]. Participants received compensation worth approximately 10 USD (either a movie ticket, two lottery scratch cards, a gift card valid in several stores, or a gift card valid in a local coffee shop). The study was reviewed by Mid Sweden University's Research Ethics Committee (January 2019), which raised no objections from an ethical point of view. All study methods and data handling were in accordance with relevant national and international regulations and recommendations.

Note that, in an initial preregistration (https://doi.org/10.17605/OSF.IO/YPHJ9), the target sample size was set by a power analysis based on the smallest key effect in Nook et al. (2015) of $d = 0.255$. In order to replicate this effect at 80%, a sample of 123 participants was necessary. However, due to the COVID-19 pandemic, data collection was suspended early after collecting data from 115 participants. Given that this provided us with 77% power to detect the smallest key effect size (which is close to the traditional threshold for power), we terminated data collection and preregistered our analyses and inclusion/exclusion rules before analyzing any data. We then excluded participants based on predetermined criteria (i.e., age > 45 years, first language not Swedish, first language English, and/or data outliers), yielding an interim sample of 105 usable participants. Preregistered analyses on this sample yielded inconclusive results: although some results allowed us to reject the null hypothesis, others yielded effect size estimates that made us concerned about Type II error due to low power. As reported in a second preregistration (https://doi.org/10.17605/OSF.IO/Z8BJK), we conducted equivalence tests (Lakens, 2017; Lakens et al., 2020) on our borderline results to investigate this concern statistically. The results of the equivalence tests indeed showed that this interim sample did not provide sufficient evidence to either affirm or reject the null hypothesis for some tests (see **Supplemental Materials**). Simulations based on interim data revealed that a sample of at least 140 usable participants would be necessary to have at least a 50% chance of finding conclusive results. We consequently preregistered a final sample size of 140 usable participants. We then collected this sample and conducted our preregistered analyses.

Procedure

After reading about the study and providing written informed consent, participants were asked to translate the three emotion words that would later be used in the task from English to Swedish to ensure that they understood the stimuli words in their L2 (all participants provided correct translations). The task used in this study was a modification of the task described in Study 1 of Nook et al. (2015). As in the original study, a trial started with a fixation cross for 500 ms, followed by a 'cue' image (i.e., a picture of a person showing a facial expression of anger, disgust, or sadness) for 1,000 ms. A blank screen then appeared for 200 ms, which was replaced by the 'target' image (either another picture of a facial expression of emotion or an emotion word) for 1,000 ms. In this study, however, the emotion word could be in English (L2; anger, disgust, or sadness) or in Swedish (L1; ilska, avsky, or sorg). The participant then had 1,000 ms to respond whether both stimuli represented the same emotion (i.e., match) or two different emotions (i.e., no-match). A blank screen was then presented for 1,500 ms before the next trial began (see **Figure 1**).

A total of 144 trials were presented in two blocks of 77 trials each, with type of trial (face, L1 word, L2 word) and congruency (congruent, incongruent) evenly distributed across the blocks. The only difference between the two blocks was the answer associated with each key so that a participant instructed to press A for 'match' and L for 'no-match' in the first block was instructed to do the reverse in the second block (initial mapping was counterbalanced across participants). Each block started with written instructions and 12 practice trials. All instructions were provided on the screen in English. The experiment was programmed and

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

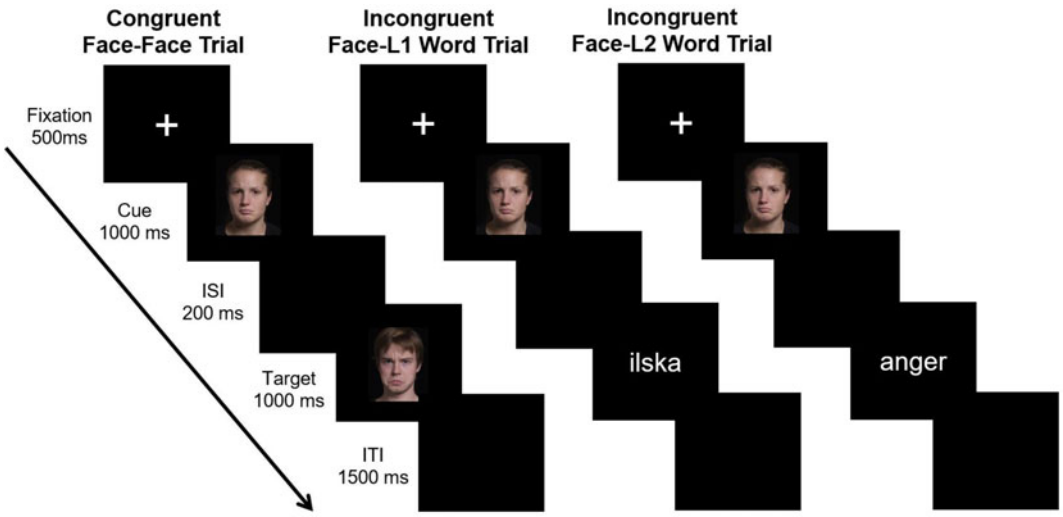


Figure 1 Study design.
Note. This figure illustrates examples of congruent and incongruent trials in the different conditions (i.e., Face, L1 Word, L2 Word). On each trial, participants first saw a fixation cross, a facial expression of emotion (called the *cue*), a blank screen inter-stimulus-interval (ISI), a facial expression of emotion or an emotion word in Swedish or English (called the *target*), and a blank screen inter-trial-interval (ITI). Participants indicated using button presses as fast as possible whether or not the emotions of the target and cue matched or did not match. Half of the trials in each condition were congruent (i.e., expressed the same emotion), and half were incongruent (i.e., expressed different emotions). This resulted in a 2 [Congruence: congruent vs. incongruent] x 3 [Context: Face, L1 word, L2 word] design assessing how each condition affects the speed and accuracy of emotion perception.

presented in MATLAB (MathWorks, 2018). Stimuli for this study were a subset of those used in the Nook et al. (2015) paper that established this paradigm. Posed anger, disgust, and sadness expressions from 63 unique individuals were selected from the NimStim (Tottenham et al., 2009) and the IASLab (Gendron et al., unpublished) validated facial stimuli sets. Three images were repeated to produce a total of 192 available stimuli. Using custom code that was run for each participant, these stimuli were pseudorandomly arranged so that there was a random arrangement of images to the 'cue' and 'target' image sets while balancing stimuli emotion and gender across all cells of the design and so that no trial included a cue and target image from the same individual. After the task, participants filled out a questionnaire with questions about their linguistic and demographic backgrounds. They were then debriefed, given their compensation, and thanked for their participation.

Instruments

LexTALE

Because studies suggest that L2-proficiency moderates the extent to which emotional resonance is dampened in a L2 (Caldwell-Harris et al., 2011; Harris, 2004), where high proficient bilinguals experience similar emotional resonance in their L1 and L2 (Degner et al., 2012; Grabovac & Pléh, 2014; Sutton et al., 2007), we included an objective measurement of proficiency in addition to self-reported proficiency. This allowed us to include both measurements as covariates in our analyses to control for the eventual effect of L2-proficiency. We used LexTALE, a standardized test to assess vocabulary knowledge in speakers of English as a second language (Lemhöfer & Broersma, 2012), as our objective L2-proficiency measurement. LexTALE correlates with other objective proficiency measurements (e.g., translation accuracy, word recognition) better than self-reported proficiency levels (Lemhöfer & Broersma, 2012). The test consists of 60 strings of letters, of which 40 are English words, and 20 are nonwords (i.e., similar to English words but not actual English words). The participant's task is to determine whether the string of letters is a word or a nonword. We followed convention (Lemhöfer & Broersma, 2012) and measured English proficiency by computing the average of the percentage of correctly identified words and the percentage of correctly rejected nonwords. Scores can range from 0 to 100, with higher scores indicating better English proficiency. In our sample, the scores ranged from 48.75 to 100 ($M = 77.48$, $SD = 11.90$) and correlated positively with self-reported L2-proficiency, $r(138) = .51$, $p < .001$, 95% CI = [.37, .62].

Data processing and analyses

Reaction times were averaged for each condition (i.e., face congruent, face incongruent, L1 word congruent, L1 word incongruent, L2 word congruent, L2 word incongruent) for every participant. Only correct responses greater than 200 milliseconds were included in the calculation of mean reaction times. To measure accuracy, signal detection methods were used (Green & Swets, 1966). Signal detection methods are recommended for psychological tasks where two types of stimuli must be discerned as they have the advantage of controlling both for detection sensitivity and response bias compared to simple percent correct accuracy measurements (Stanislaw & Todorov, 1999). Again, non-responses and responses faster than 200 milliseconds were excluded from computations. Sensitivity (d') was defined as z (hit rate) – z(false alarm rate) and was calculated for each condition. The inclusion and exclusion criteria for reaction time and sensitivity presented above were the same as in Nook et al. (2015).

Analyses follow our preregistered plan except where noted. We first report analyses of reaction time and sensitivity across all conditions to give an overall assessment of how these conditions differed. We analyzed reaction time using a 2 [Congruence: congruent vs. incongruent] x 3 [Context: face vs. L1 word vs. L2 word] repeated-measures ANOVA, and we analyzed sensitivity

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

using a one-way repeated-measures ANOVA. To answer our first research question, where we hypothesized that the effect of emotion words found in Nook et al. (2015) would replicate in a Swedish population, we conducted a more focused 2 [Congruence: congruent vs. incongruent] x 2 [Context: face vs. L1 word] repeated-measures ANOVA of reaction times. We hypothesized a significant interaction: although responses to congruent trials should overall be faster than incongruent trials, we expected this to be especially pronounced in the L1 context. We then tested whether sensitivity results would also replicate prior research using a paired-sample *t*-test comparing face and L1 contexts. We hypothesized that sensitivity would be higher in the L1 context than in the face context.

To answer our second question and examine whether L1 words are more strongly associated with emotion concepts than L2 words, we analyzed reaction times with a focused 2 [Congruence: congruent vs. incongruent] x 2 [Context: L1 word vs. L2 word] repeated-measures ANOVA. We again hypothesized an interaction such that L1 contexts would speed congruent trial responding more than L2 contexts. We also compared sensitivity in L1 and L2 contexts using a paired-sample *t*-test and hypothesized higher sensitivity in the L1 than L2 context. Finally, to answer our last research question regarding moderation by proficiency, we conducted an ANCOVA by adding self-reported proficiency and LexTALE scores as additional predictors in the analysis described above. We centered continuous predictors prior to analyses. We hypothesized that we would observe significant interactions indicating that the effects described above would be intensified for Swedish speakers who are more proficient in English.

We provide effect sizes and confidence intervals (CIs) for all analyses. For *t*-tests, we report Cohen's *d* and 95% CIs. For ANOVAs, we provide partial eta-squared (η_p^2) and 90% CIs due to the fact that tests of the F distribution are single-tailed, meaning that the 90% CI gives the range over which we are 95% confident we can reject the null hypothesis (Lakens, 2013).

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Data and analytic code for this study are available here (https://osf.io/ud5f3/). This study was preregistered before any analyses were conducted (https://doi.org/10.17605/OSF.IO/YPHJ9), and the preregistration was amended following the conclusion that further data were required (https://doi.org/10.17605/OSF.IO/Z8BJK). For full transparency, we provide results of analyses from the initial preregistration in the **Supplemental Materials**. How we determined our sample size, all data exclusions, all manipulations, and all measures in the study are reported in the Method above.

Results

Overall analyses

As preregistered, we first report analyses of reaction time and sensitivity from all conditions before following up main effects and potential interactions with focused tests of key research questions. A 2 [Congruence: congruent vs. incongruent] x 3 [Context: face vs. L1 word vs. L2 word] repeated-measures ANOVA showed a significant main effect of Congruence, $F(1, 139) = 259.17$, $p < .001$, $\eta_p^2 = .65$, 90% CI = [.57, .71], a main effect of Context, $F(2, 278) = 438.74$, $p < .001$, $\eta_p^2 = .76$, 90% CI = [.72, .79], and a trending interaction between Congruence and Context, $F(2, 278) = 2.97$, $p = .053$, $\eta_p^2 = .02$, 90% CI = [0, .05] (**Figure 2a**). A one-way repeated-measures ANOVA of sensitivity also revealed a significant main effect of Context, $F(2, 278) = 46.02$, $p < .001$, $\eta_p^2 = .25$, 90% CI = [.18, .31] (**Figure 2b**). As such, we found differences in RT and sensitivity when examining all conditions, motivating focused follow-up analyses.

Replicating prior work: does language facilitate emotion perception?

We next conducted an analysis that was parallel to that of Nook et al. (2015) in which we ignored L2 word trials and conducted a 2 [Congruence: congruent vs. incongruent] x 2 [Context: face vs. L1 word] ANOVA on RT. This analysis asks whether first language emotion words facilitate emotion perception more than facial expressions, even in native Swedish speakers, providing a test to ensure that the paradigm was valid in another population with a different language. As hypothesized, we found a main effect of Congruence, $F(1, 139) = 195.21$, $p < .001$, $\eta_p^2 = .58$, 90% CI = [.50, .65], such that responses were faster for congruent trials ($M = 0.68$s, $SD = 0.06$) than incongruent trials ($M = 0.73$s, $SD = 0.05$). This supports the notion that cues were more strongly associated with targets when cue and target shared an emotion concept (see Introduction). We also observed a main effect of Context, $F(1, 139) = 461.35$, $p < .001$, $\eta_p^2 = .77$, 90% CI = [.71, .81], such that RTs were faster for L1 word targets ($M = 0.65$s, $SD = 0.06$) than for face targets ($M = 0.76$s, $SD = 0.05$). Finally, a significant interaction between Congruence and Context emerged, $F(1, 139) = 5.71$, $p = .018$, $\eta_p^2 = .04$, 90% CI = [.004, .10]. Follow-up paired-sample *t*-tests revealed that the impact of congruence on L1 word trials, $t(139) = 12.87$, $p < .001$, $d = 0.79$, 95% CI = [0.65, 0.93], was larger than the impact of congruence on face trials $t(139) = 8.53$, $p < .001$, $d = 0.72$, 95% CI = [0.53, 0.90]. This result replicates prior work, showing that facial expressions are associated with first language emotion words more strongly than other facial expressions.

Regarding sensitivity (d'), a significant paired-sample *t*-test, $t(139) = 8.52$, $p < .001$, $d = 0.77$, 95% CI = [0.57, 0.98], showed that participants were significantly more sensitive (i.e., had stronger abilities to accurately match cues and targets) on L1 word trials ($M = 1.27$, $SD = 0.59$) than face trials ($M = 0.85$, $SD = 0.47$). This again replicates prior research that d' is higher when matching facial expressions to first language emotion words than when matching with other facial expressions.

Comparing emotion words in first vs. second language

To specifically compare L1 and L2 contexts for emotion perception (in order to test whether this could be an underlying mechanism that would explain lower L2 emotional resonance), we then ignored face trials and analyzed RT using a 2 [Congruence: congruent vs. incongruent] x 2 [Context: L1 word vs. L2 word] ANOVA. We again found a main effect of Congruence, $F(1, 139) = 226.84$, $p < .001$, $\eta_p^2 = .62$, 90% CI = [.54, .68], such that responses were faster for congruent trials ($M = 0.62$s, $SD = 0.07$) than incongruent trials ($M = 0.67$s, $SD = .06$). However, we did not observe a main effect of Context for L1 and L2 words ($M = 0.65$s, $SD = 0.06$), $F(1, 139) = 0.44$, $p = .507$, $\eta_p^2 = .003$, 90% CI = [0, .04], nor an interaction between Congruence and Context, $F(1, 139) = 2.44$, $p = .121$, $\eta_p^2 = .02$, 90% CI = [0, .07]. A follow-up

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

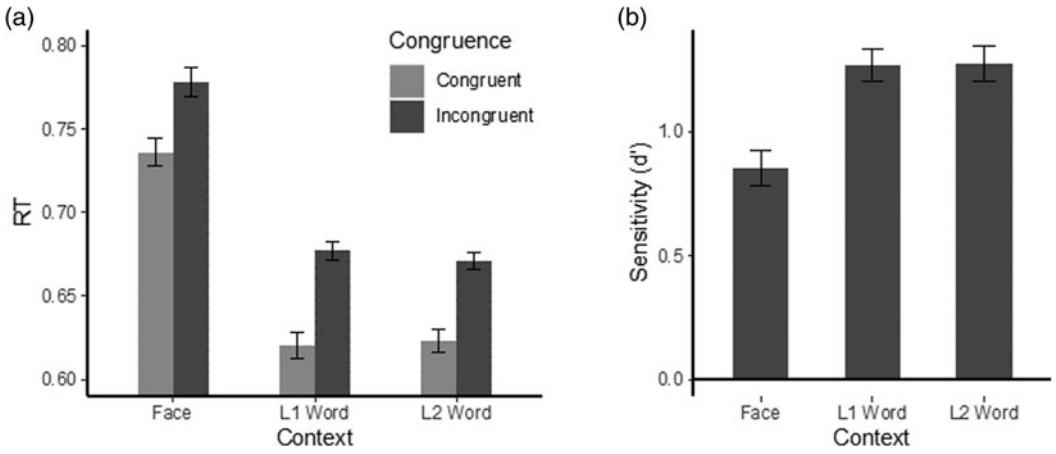


Figure 2 Overall Reaction Times and Accuracy

Note. Panel A displays reaction times (RT) for each condition. Main effects indicated that participants were faster for congruent than incongruent trials and for both L1 and L2 word trials than face trials. Preregistered ANOVAs showed that the difference between congruent and incongruent reaction times was smaller for face trials and L1 word trials, indicating that emotion faces prime congruent L1 emotion words more than congruent emotion facial expressions, replicating prior work. Contrary to hypotheses, however, the impact of congruence on L1 and L2 words was similar, indicating similar effects for both L1 and L2 emotion words. Panel B displays sensitivity (d'), a signal detection measure of accuracy, for each condition. Participants were less able to correctly identify cue-target matches in the face condition than both the L1 and L2 word conditions. Error bars represent 95% confidence intervals adjusted for within-subject comparisons.

paired-samples *t*-test revealed that the impact of congruence on L2 word trials, $t(139) = 10.61$, $p < .001$, $d = 0.70$, 95% CI = [0.55, 0.84], was not significantly different than for L1 word trials ($d = 0.79$, see above). Thus, contrary to hypotheses, the extent to which facial expressions of emotion are associated with L1 and L2 emotion words did not significantly differ in this study.

Similarly, contrary to hypotheses, a paired-sample t-test showed a non-significant difference, $t(139) = 0.15$, $p = .879$, $d = 0.01$, 95% CI = [-0.14, 0.17], between d' for L1 trials ($M = 1.27$, $SD = 0.59$) and L2 trials ($M = 1.27$, $SD = 0.62$). Sensitivity was also significantly higher in L2 trials than Face trials, $t(139) = 7.80$, $p < .001$, $d = 0.76$, 95% CI = [0.54, 0.97]. Thus, native Swedish speakers were not significantly different in their ability to match facial expressions of emotion with L1 vs. L2 emotion words.

Influence of proficiency

To test the potential effect of L2-proficiency on task behavior, we used both self-reported English proficiency scores and LexTALE scores as additional moderating factors in the analyses described above. We then unpacked significant effects and interactions with follow-up analyses.

First, a 2 [Congruence: congruent vs. incongruent] x 3 [Context: faces vs. L1 word vs. L2 word] repeated-measures x self-reported English proficiency ANCOVA was used to analyze RT (**Figure 3a**). We observed main effects of Congruence, $F(1, 138) = 243.35$, $p < .001$, $\eta_p^2 = .65$, 90% CI = [.57, .71] and Context, $F(2, 276) = 470.62$, $p < .001$, $\eta_p^2 = .77$, 90% CI = [.74, .80]. However, we also observed a significant main effect of self-reported L2-proficiency, $F(1, 138) = 4.12$, $p = .044$, $\eta_p^2 = .03$, 90% CI = [.0004, .09], which was qualified by a significant interaction between L2-proficiency and Context, $F(1, 138) = 11.10$, $p < .001$, $\eta_p^2 = .07$, 90% CI = [.03, .12]. No other interactions were significant, $ps > .3$. We then unpacked the L2-proficiency main effect and interaction by computing correlations between L2-proficiency and average RT for each Context. We observed a significant negative relationship between L2-proficiency and average RT in both the L1, $r(138) = -.23$, $p = .006$, and L2 contexts, $r(138) = -.23$, $p = .007$. However, L2-proficiency did not correlate with RT in the face context, $r(138) = .05$, $p = .585$. As such, greater self-reported L2-proficiency was associated with faster RTs to both L1 and L2 trials but not face trials.

Results for the ANCOVA including LexTALE scores were similar but differed slightly (**Figure 3b**). We observed main effects of both Congruence, $F(1, 138) = 275.69$, $p < .001$, $\eta_p^2 = .67$, 90% CI = [.59, .72], and Context, $F(2, 276) = 440.57$, $p < .001$, $\eta_p^2 = .76$, 90% CI = [.72, .79]. Although no main effect of LexTALE scores emerged, $F(1, 138) = 0.19$, $p = .668$, $\eta_p^2 = .001$, 90% CI = [0, .03], LexTALE scores significantly interacted with Congruence, $F(1, 138) = 9.86$, $p = .002$, $\eta_p^2 = .07$, 90% CI = [.02, .14]. No other interactions were significant, $ps > .05$. To unpack this interaction, we computed participants' 'association' scores for each context by subtracting their average RT on Congruent trials from their average RT on Incongruent trials within each context. We then tested how LexTALE scores correlated with these association scores. LexTALE scores correlated negatively with association scores on Face trials, $r(138) = -.30$, $p < .001$, but not L1 word trials, $r(138) = -.11$, $p = .193$, or L2 word trials, $r(138) = -.08$, $p = .325$. As such, greater objectively measured L2-proficiency was associated with less congruent cue association on face trials.

Turning to d', we first conducted a 3 [Context: congruent vs. incongruent] x self-reported L2-proficiency ANCOVA (**Figure 3c**). Interestingly, we again observed a main effect of Context, $F(1, 276) = 46.10$, $p < .001$, $\eta_p^2 = .25$, 90% CI = [.18, .31]. However, we also observed a significant main effect of L2-proficiency, $F(1, 138) = 10.13$, $p = .002$, $\eta_p^2 = .07$, 90% CI = [.02, .14]. No L2-proficiency x Context interaction emerged, $p = .283$. Results for the LexTALE analysis were the same (**Figure 3d**), revealing only main effects of Context, $F(1, 276) = 46.00$, $p < .001$, $\eta_p^2 = .25$, 90% CI = [.18, .31], and LexTALE, $F(1, 138) = 5.56$, $p = .020$, $\eta_p^2 = .04$, 90% CI = [.003, .10], and no interaction, $p = .379$. Thus, participants with higher L2-proficiency (both self-reported and objectively assessed) showed higher performance on this task. Although figures suggest that this

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

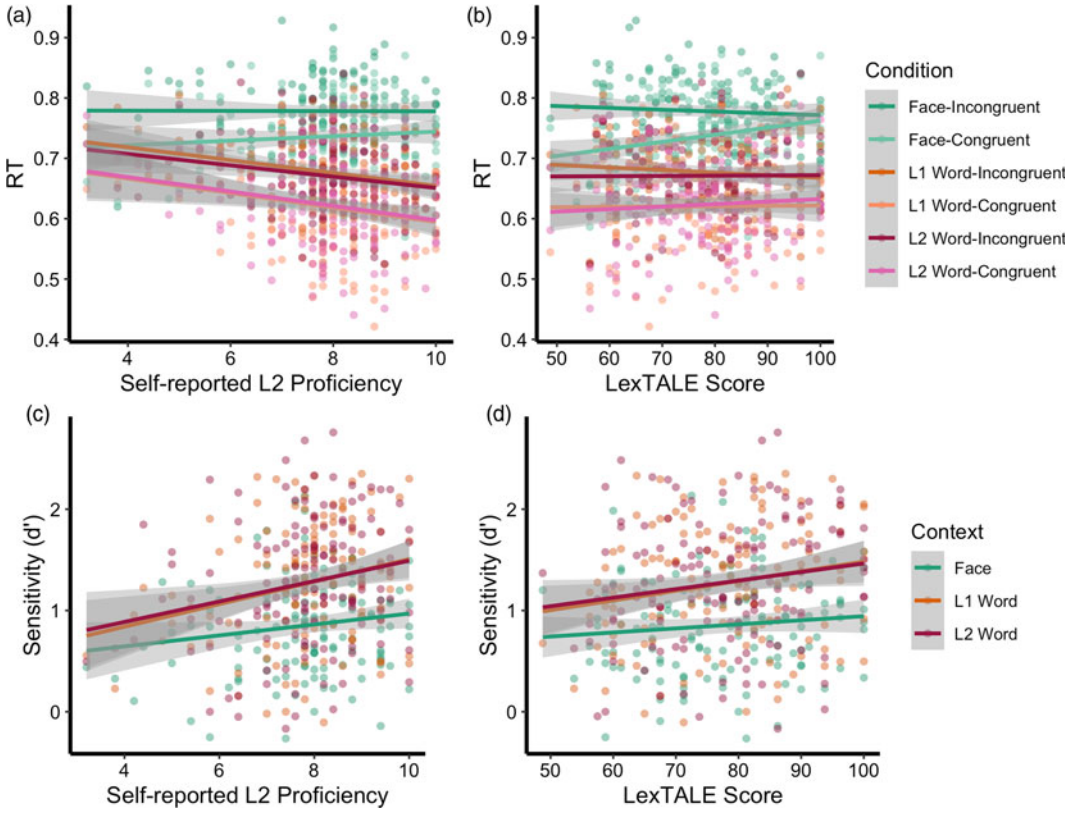


Figure 3 Relations with Self-Reported L2-proficiency and LexTALE Scores

Note. Panel A presents relations between reaction time (RT) and self-reported L2-proficiency (non-centered scores) for each condition. Preregistered analyses show that higher L2-proficiency is related to faster responding for congruent and incongruent trials in L1 and L2 word trials (dark orange, light orange, dark red, and pink lines). Panel B represents relations between RT and objectively assessed L2-proficiency (non-centered scores) using the LexTALE test. Results show that increasing LexTALE scores correlated with reductions in the difference between Face-Incongruent (light green) and Face-Congruent (dark green) trials. As such, individuals with lower L2-proficiency showed greater tendency for facial expressions to speed responding of congruent facial expressions. Panels C and D present relations between sensitivity (d'), a signal-detection measure of accuracy, and both self-reported and LexTALE L2-proficiency for face (green), L1 (orange), and L2 word (red) trials. Results for both analyses show that increased L2-proficiency is related to higher overall sensitivity. No interaction with context conditions emerged. Grey-shaded regions represent 95% confidence intervals.

relationship was specific to word conditions, interactions were not significant.

Discussion

In this study, we investigated whether the emotion concepts are more strongly associated with emotion words in a first language rather than a second language in a population of Swedish–English bilinguals with Swedish as their first language (L1) and English as a second language (L2). As in Nook et al. (2015), we asked our participants to indicate whether an emotion word or a face represented the same emotion as a facial expression of an emotion that was presented immediately before. In this study however, the target emotion word could be in Swedish (L1) or in English (L2). We hypothesized that participants would be faster and more accurate in the L1 word condition compared to both the L2 word condition and to the face condition, and that they would also be faster and more accurate in the L2 word condition compared to the face condition. We also tested whether RT analyses would indicate that congruent cue-target pairs would more strongly speed responses to L1 words than L2 words and faces, providing evidence for a stronger association between concepts and L1 words than these other targets. Finally, we tested whether self-reported proficiency in English and an objective measurement of English proficiency (LexTALE) would moderate these relationships. We found evidence in line with some but not all of these hypotheses, ultimately suggesting that (i) both L1 and L2 emotion words speed and facilitate emotion perception compared to facial expressions of emotion; (ii) behavior in response to L1 and L2 emotion words did not significantly differ in this study; and (iii) higher L2-proficiency correlated with increased speed in word conditions, lower 'association' scores in the face condition, and increased accuracy across all conditions.

As hypothesized, we found that participants were faster and more accurate in both the L1 and L2 word condition compared to the face condition, and they were faster when responding to congruent trials compared to incongruent trials in all conditions. There was also a significant interaction effect between condition and congruence when we compared the reaction times in the L1 word and face conditions, such that the effect of congruence was larger for the L1 word condition than for the face condition. As such, our results replicate the findings from Nook et al. (2015) but using another language and set in another culture (i.e., Sweden). An additional replication of these results in such a dissimilar setting provides increased confidence in these results and lends credence to the stability of the paradigm.

Contrary to our hypotheses, though, participants' performance did not differ between the L1 and L2 word conditions.

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

Specifically, their responses did not significantly differ in their speed and accuracy in both conditions, and we did not find that congruence in cue-target emotions sped responding more for L1 vs. L2 emotion words. This suggests that in this sample both L1 and L2 emotion words facilitated emotion perception but that they did so to similar degrees. This contrasts with our hypotheses, potentially suggesting that decreased emotional resonance in L2 settings may not be attributable to decreased associations between L2 emotion words and emotion concepts. In line with this notion, neuroimaging studies suggest that there is strong overlap in neural activation by L1 and L2 words in bilinguals (e.g., Marian et al., 2003; Martin et al., 2009; Rodriguez-Fornells et al., 2002). As such, it could be the case that understanding an L2 emotion word and knowing its L1 translation is sufficient to produce indistinguishable behavior on tasks like these. Alternatively, it is important to note that we provide here just a single test of this general question, and additional studies are needed to reach the conclusion that emotion concepts are similarly associated with L1 and L2 emotion words. Because we did not measure actual emotional resonance in this sample, it is not obvious that this group of participants would actually experience reduced emotional reactions in L2 settings. As such, it is unclear whether we would find this null result in different settings or in samples that specifically include participants who experience diminished emotional resonance. Indeed, it may be that our sample's L2-proficiency in English was sufficiently high to create an efficient context for emotion perception that was similar to the L1 context. Although English was a second language for all our participants and that they reported being significantly less proficient in English compared to Swedish, their self-reported proficiency and LexTALE scores were both high. This aligns with the literature summarized in the introduction showing that highly proficient bilinguals experienced similar levels of emotional resonance in both their L1 and in their L2, and with Sutton et al. (2007) who failed to find a difference between highly proficient bilinguals' L1 and L2 on an emotional Stroop task. To disentangle these competing hypotheses, studies where participants have a wider range of L2-proficiency will be necessary, a point we return to below.

Our own L2-proficiency results provide some support for this notion. We found that the more proficient participants reported being in English, the faster they responded in the L1 word and L2 word conditions. LexTALE scores were not associated with overall speed; instead, higher performance on this test correlated with reduced congruent-cue speeding in the face condition only. Finally, both higher self-reported proficiency and higher LexTALE scores were associated with higher accuracy across all conditions. Although self-reported L2-proficiency and LexTALE interacted with different aspects of the task (i.e., condition and congruence respectively), they both point in the same direction. Specifically, English proficiency correlated with better and faster emotion perception. While it is not clear why an effect with self-reported L2-proficiency was found in both L1 and L2 word conditions, a possibility is that self-reported L2-proficiency is a proxy for, or correlates highly with, linguistic skills in general. Although a significant interaction with LexTALE scores was only found in the face condition, it is still in line with the interaction found for self-reported proficiency and might simply show the other side of the coin. Namely, lower LexTALE scores correlated with more priming in the face condition; in other words, the facilitating effect of non-linguistic cues (i.e., faces) was GREATER in people with LOWER L2-proficiency. High priming in face-face trials indicates that responses are faster when the physical facial movements of the cue and target match (e.g., two frowns, rather than a frown followed by a grimace), which suggests that attention is focused on non-linguistic features of the task. This approach seems heightened in people with low L2-proficiency, who also showed reduced accuracy (as measured by d') across all conditions of the task.

Viewed this way, our proficiency results are conceptually consistent: higher language proficiency correlates with greater speed in word conditions and less facilitation from non-linguistic physical features of the faces. In fact, 'bilingualism' (as measured by improved proficiency in a second language) correlates with better emotion perception, as the more proficient a speaker is in a L2, the more accurately they perceived expressions of emotion. However, before calling this effect for (yet) another bilingual advantage (for a discussion of the problematic use of the terms 'bilingual advantage' and 'bilingualism' in research during the last few decades, see Luk, 2022), it is critical to consider that higher L2-proficiency may also be a proxy for other factors that can have affected performance on the task, such as better cognitive functions or higher IQ. Therefore, until the underpinnings of the effects that we found in this study can be explained – and until the mechanisms that underlie the differences in executive functions between monolinguals and bilinguals (if they exist) are understood – it is still too early to make this interpretation definitively. Still, this finding suggests that LANGUAGE (not only a first language) should be taken seriously as a factor in accurate emotion perception.

Critically, the generalizability of our results to other cultural contexts and/or language pairs should be made with caution as our results may only reflect the particular linguistic context and high-proficiency sample studied here. Indeed, there is a growing awareness in the field of bilingualism research for the importance of considering the influence that socio-cultural factors may have on all aspects of bilingualism (Titone & Tiv, 2023). According to this view, effects found in a specific bilingual population in a particular societal and historical context cannot be automatically assumed to be identical or even similar in another bilingual population, especially if they are in a different context. For instance, the differences in emotion concepts in Swedish and English, which both are Indo-European Germanic languages predominantly found in western cultures, are arguably smaller than between Swedish and say a Sino-Tibetan language such as Chinese. Additionally, in the specific case of English in Sweden, it is a language that is taught early in school systems and widely used in the society. Indeed, for many Swedes, English is a language that they come in contact with on a daily or almost daily basis through ads, games, and a variety of media, art, and entertainment such as television series, movies, and music (Institutet för språk och folkminne, 2021). Relatedly, a recent study investigating social categorization based on language found that Swedish participants remember English and Swedish speakers equally well in a surprise memory task (suggesting an in-group categorization for both types of speakers), but remember speakers of other languages (i.e., Spanish) more poorly than Swedish speakers (Champoux-Larsson et al., 2022). This strengthens the idea that English holds a special place in status in the Swedish society.

Therefore, our results may be limited to a second language that is widespread in the population and where proficiency is generally high, and also where the nature of the encounters with the L2 is often emotionally loaded. For these reasons, future research should investigate other language pairs where emotion words

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

and concepts do not overlap as much as they arguably do in Swedish and English, where more variability in L2-proficiency can be found, and where the L2 is not a highly emotional language that is an integral part of the L1 society. It would also be informative to explore languages and emotion words where translations between languages are non-existent (e.g., the word 'saudade' in Portuguese) or when the translations overlap poorly (e.g., the meaning of the word 'revnost' in Russian which differs from the meaning of its typical English translation, 'jealousy'). Furthermore, although we expected to find a difference between the L1 word and L2 word conditions based on the body of research suggesting lower L2 emotional resonance in combination with constructionist accounts of emotion (according to which the processes that lead us to construct and understand both what we are feeling and what OTHER people are feeling are similar), it is possible that the paradigm that we used did not tap into the mechanisms of emotional resonance. For instance, studies using memory tasks to investigate emotional resonance in a L1 and a L2 (e.g., Ayçiçegi & Harris, 2004; Ayçiçegi-Dinn & Caldwell-Harris, 2009; Ferré et al., 2010) also failed to find a difference between the two language contexts. To address this limitation, using a similar task as we used here but that involves actual experiences may be more appropriate to test perceived emotional resonance specifically. Additionally, it may be that behavioral differences could not be detected with our task, but that psychophysiological measurements might show another pattern. For instance, Eilola and Havelka (2011) did not find behavioral differences between L1 and L2 emotional words in a Stroop task but found psychophysiological differences (skin conductance) suggesting lower L2 emotional resonance. This suggests that, depending on the task, physiological responses may differ in L1 and L2 contexts without being detectable behaviorally.

Constraints on generality and limitations

Our sample included adults aged 18-45 in Sweden who spoke Swedish as a first language and English as a second language. The sample was a majority female, but educational levels were largely representative of the general population. Collecting information on ethnicity and race was not allowed due to Swedish ethics policies. That said, we advertised for the study in a wide variety of settings and aimed to ensure a diverse sample within our selection criteria. Nonetheless, there are several important limits to confidence based on these constraints that should be kept in mind. As we discuss above, our results may not necessarily apply to other cultural contexts and/or language pairs, especially in populations where L2-proficiency is lower. Because this study was the first to investigate visual perception of emotion in relation to first and second languages specifically, more research will be needed to establish the extent to which findings hold in other contexts. We also cannot draw inferences about youths or older adults from this study, given the constrained age range. Additionally, the higher proportion of women to men and the lack of information on race or ethnicity indicates that future research should attend to these factors when replicating and extending these findings. In addition to age, gender and education, other important factors that future research should consider are measures of general intelligence, cognitive functioning, and general linguistic skills. Given that these factors all are relevant for the type of task that was used in this study, it will be important to determine how they potentially interact with performance and conceptual priming.

Conclusion

Overall, this study was the first to investigate the effect of a second language on emotion perception. In line with constructionist accounts of emotion perception, our results show that language creates a facilitating context for emotion perception, even when another language than English is used, and that emotion perception is not impaired when the linguistic context is in a second, and perhaps less emotionally intense, context. However, more research is needed to investigate the effects that we found in other cultures and with other language pairs.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728923000998

Acknowledgements. The authors would like to thank Alice Eliasson, Amani Lavefjord, and Lina Andersson for their assistance with data collection. This work was supported by a National Science Foundation Graduate Research Fellowship (DGE1144152) and Graduate Research Opportunities Worldwide (GROW) Fellowship to E.C.N. This study's design, hypotheses, and analyses were preregistered; see https://osf.io/ud5f3/registrations. Data and analytic code for this study are available here: https://osf.io/ud5f3/. The authors have no competing interests to declare.

References

Ayçiçegi, A., & Harris, C. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *18*(7), 977–987. https://doi.org/10.1080/02699930341000301

Ayçiçegi-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach*. *Bilingualism: Language and Cognition*, *12*(03), 291. https://doi.org/10.1017/S1366728909990125

Barrett, L. F. (2006). Solving the Emotion Paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*(3), 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Caldwell-Harris, C. L., Tong, J., Lung, W., & Poo, S. (2011). Physiological reactivity to emotional phrases in Mandarin–English bilinguals. *International Journal of Bilingualism*, *15*(3), 329–352. https://doi.org/10.1177/1367006910379262

Champoux-Larsson, M.-F., Ramström, F., Costa, A., & Baus, C. (2022). Social categorization based on language and facial recognition. *Journal of Language and Social Psychology*, *41*(3), 331–349. https://doi.org/10.1177/0261927X211035159

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *BILINGUALISM-LANGUAGE AND COGNITION*, *15*(1), 181–189. https://doi.org/10.1017/S1366728911000095

Der, G., & Deary, I. J. (2009). "Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey": Correction. *Psychology and Aging*, *24*(1), 229–229. https://doi.org/10.1037/a0015515

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, *25*(2–3), 204–222. https://doi.org/10.1080/01434630408666529

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. (2010). *Emotions in multiple languages*. Palgrave Macmillan.

Doyle, C. M., Gendron, M., & Lindquist, K. A. (2021). Language is a unique context for emotion perception. *Affective Science*, *2*(2), 171–177. https://doi.org/10.1007/s42761-020-00025-7

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*(1), 62–73. https://doi.org/10.1037/xge0000361

Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, *15*(3), 353–369. https://doi.org/10.1177/1367006910379263

Ferré, P., García, T., Fraga, I., Sánchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition & Emotion*, *24*(5), 760–785. https://doi.org/10.1080/02699930902985779

Ferré, P., Guasch, M., Stadthagen-Gonzalez, H., & Comesaña, M. (2022). Love me in L1, but hate me in L2: How native speakers and bilinguals rate the affectivity of words when feeling or thinking about them. *Bilingualism: Language and Cognition*, *25*(5), 786–800. https://doi.org/10.1017/S1366728922000189

Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: Evidence for the role of verbal labels in categorical perception of emotion. *Emotion*, *10*(4), 544–554. https://doi.org/10.1037/a0019017

Gendron, M., Lindquist, K. A., & Barrett, L. F. (unpublished). *Ratings of IASLab facial expression stimuli*. http://www.affective-science.org/

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*(2), 314–325. https://doi.org/10.1037/a0026007

Grabovac, B., & Pléh, C. (2014). Emotional activation measured using the emotional Stroop task in early Hungarian-Serbian bilinguals from Serbia. *Acta Linguistica Hungarica*, *61*(4), 423–441. https://doi.org/10.1556/ALing.61.2014.4.3

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.

Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development*, *25*(2–3), 223–247. https://doi.org/10.1080/01434630408666530

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, *24*(4), 561–579. https://doi.org/10.1017/S0142716403000286

Heredia, R. R., & Cieślicka, A. B. (2014). Bilingual memory storage: Compound-coordinate and derivatives. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of Bilingual Memory* (pp. 11–39). Springer New York. https://doi.org/10.1007/978-1-4614-9218-4_2

Institutet för språk och folkminne. (2021). *Engelskan i Sverige [English in Sweden]*. https://www.isof.se/lar-dig-mer/kunskapsbanker/lar-dig-mer-om-flersprakighet/engelskan-i-sverige

Kovelman, I., Baker, S. A., & Petitto, L.-A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, *11*(2), 203–223. https://doi.org/10.1017/S1366728908003386

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology: Series B*, *75*(1), 45–57. https://doi.org/10.1093/geronb/gby065

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

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science*, *19*(9), 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, *6*(1), 125–138. https://doi.org/10.1037/1528-3542.6.1.125

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66–71. https://doi.org/10.1177/1754073912451351

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, *14*(2), 375–387. https://doi.org/10.1037/a0035293

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do More than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99–108. https://doi.org/10.1177/0963721414553440

Luk, G. (2022). Justice and equity for whom? Reframing research on the "bilingual (dis)advantage." *Applied Psycholinguistics*, 1–15. https://doi.org/10.1017/S0142716422000339

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7

Martin, C. D., Dering, B., Thomas, E. M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *NeuroImage*, *47*(1), 326–333. https://doi.org/10.1016/j.neuroimage.2009.04.025

MathWorks. (2018). *MATLAB* (Version R2018b) [Computer software]. MathWorks.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic Processes in Reading: Visual Word Recognition* (pp. 264–336). Taylor & Francis Group.

Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion*, *15*(5), 569–578. https://doi.org/10.1037/a0039166

Pavlenko, A. (2005). *Emotions and Multilingualism*. Cambridge University Press.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, *11*(2), 147–164. https://doi.org/10.1017/S1366728908003283

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nösselt, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature*, *415*(6875), 1026–1029. https://doi.org/10.1038/4151026a

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, *32*(4), 863–881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Satpute, A. B., Nook, E. C., Narayanan, S., Shu, J., Weber, J., & Ochsner, K. N. (2016). Emotions in "black and white" or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. *Psychological Science*, *27*(11), 1428–1442. https://doi.org/10.1177/0956797616661555

Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion*, *30*(3), 488–500. https://doi.org/10.1080/02699931.2015.1018144

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion*, *21*(5), 1077–1090. https://doi.org/10.1080/02699930601054133

Thoma, D., Hüsam, J., & Wielscher, K. (2023). Introducing grip force as a nonverbal measure of bilingual feelings. *Bilingualism: Language and Cognition*, *26*(1), 125–137. https://doi.org/10.1017/S1366728922000396

Titone, D. A., & Tiv, M. (2023). Rethinking multilingual experience through a systems framework of bilingualism. *Bilingualism: Language and Cognition*, *26*(1), 1–16. https://doi.org/10.1017/S1366728921001127

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press

Toivo, W., & Scheepers, C. (2019). Pupillary responses to affective words in bilinguals' first versus second language. *PLOS ONE*, *14*(4), e0210450. https://doi.org/10.1371/journal.pone.0210450

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Weimer, A., Koniakowsky, I., Nazir, T. A., & Huckauf, A. (2022). Behavioral evidence for differences in emotional resonance during processing first and second language. *International Journal of Bilingualism*, *26*(6), 784–798. https://doi.org/10.1177/13670069211073578

Winskel, H. (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai-English bilinguals. *INTERNATIONAL JOURNAL OF PSYCHOLOGY*, *48*(6), 1090–1098. https://doi.org/10.1080/00207594.2013.793800

https://doi.org/10.1017/S1366728923000998 Published online by Cambridge University Press