

Language-dependent emotions in heritage and second language bilinguals: When physiological reactions deviate from feelings

International Journal of Bilingualism
2024, Vol. 28(3) 374–389
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13670069231159840
journals.sagepub.com/home/ijb



Dieter Thoma 
University of Mannheim, Germany

Abstract

Aims and objectives: Whether bilinguals show language-dependent emotions often depends on the emotion measure used. Here, we examine if differences between automatic pupil reactions and self-reported feelings in response to an emotional narrative presented in a first, second, or heritage language (HL) indicate different stages of emotion processing.

Methodology: German HL speakers of Russian and Turkish ($n = 72$) and German second language (L2) speakers of English and French ($n = 89$) listened to a video-based emotional narrative in German or their other language and rated how they felt about it (arousal and valence). We contrasted pupil diameter during the video with a language-specific baseline. Age of acquisition, language use frequency in emotional contexts, and language proficiency were used to verify that HL speakers were balanced simultaneous and L2 speakers unbalanced sequential bilinguals.

Data and analysis: Linear mixed-effects models were fitted to the pupillometry data and ordinal logistic models to the self-report data.

Findings and conclusions: HL speakers showed similar automatic reactions in both languages but rated the German narrative less emotional. L2 speakers showed weaker automatic reactions in L2 yet rated the narrative similar in both languages. This reversed pattern confirmed that automatic and conscious emotion measures tap into different stages of bilingual emotion processing. Furthermore, language-dependent emotions in self-reports seem to be linked to sociocultural frames that go beyond the scope of context and processing-based explanations.

Originality: The study is among the first to systematically examine discrepancies between automatic and conscious measures of bilingual language-dependent emotions with different types of bilinguals and within one experimental paradigm.

Significance: The findings imply that theories of bilingual emotions need further development to explain consistently and explicitly why language-dependent emotional reactions vary with bilingualism and emotion measures. Methodologically, the findings advocate for multi-measure approaches to enhance the validity of future research.

Keywords

Bilingualism, emotion, feelings, pupillometry, self-report

Corresponding author:

Dieter Thoma, Department of English, Psycholinguistics, University of Mannheim, Schloss, Mannheim 68131, Germany.
Email: thoma@uni-mannheim.de

Introduction

Bilinguals' emotional reactivity can depend on the language used to verbalize emotional situations. This observation has been conceptualized against the background that language co-constructs emotions by providing labels as "happiness," "anger," and "fear" to categorize and communicate diffuse affectual sensations as emotional experiences (Lindquist et al., 2015). The emotional outcome of linguistic co-construction can vary because languages differ in their emotion mapping and vocabulary (e.g., Jackson et al., 2019), and most bilinguals' lexical knowledge, processing skills, and language use are not equivalent in their languages. Empirically, however, the evidence for language-dependent emotions varies across emotion measures (see Table 1). Automatic, neurophysiological, and cognitive measures of emotion processing provide relatively consistent support for differential and usually weaker reactions to emotion stimuli verbalized in the weaker second language (L2), in contrast to the dominant and more proficient first language (L1)—at least for negative or intense stimuli. In contrast, conscious self-reports of perceived emotionality along basic affectual dimensions (arousal and valence) typically suggest similar reactions in L1 and L2. Table 1 summarizes the key characteristics of 11 multi-measure studies that combined neurophysiological or cognitive measures with (meta-cognitive) self-report measures of emotion. Event-related potentials, electrodermal skin conductance responses, pupillometry, and lexical processing mostly confirmed either a main effect of language or a language–emotionality interaction, such that reactions were weaker in L2 or the differences between stimuli of variable arousal/valence were less pronounced in L2, relative to L1. Yet, self-reported feelings on verbal or emoticon scales showed no or opposite effects of language within the same experiments.

The lack of alignment of findings derived from different emotion measures is a common observation in emotion research (Mauss et al., 2005; Mauss & Robinson, 2009), and it has been recognized in the earliest multi-measure approaches to bilingual emotions (e.g., Anooshian & Hertel, 1994). An additional challenge in bilingual emotion research is that self-reported feelings in multi-measure studies often do not differ between languages (see Table 1, last column), whereas single-measure self-report studies typically find that L1 feels "more emotional" than L2 (e.g., Dewaele, 2016; Garrido & Prada, 2021; Pavlenko, 2012; Puntoni et al., 2009). Therefore, it is unclear if the measurement inconsistency represents methodological issues or conceptual differences.

Methodologically, several authors have argued that self-reports of emotion are unreliable, in particular, if they include language themselves (de Langhe et al., 2011; Harris et al., 2006). Furthermore, inter-experimental comparisons are difficult because most studies differ in the types of bilinguals sampled and the emotion tasks used (see Table 1). Several studies tested immigrant populations who lived in an L2 environment and sometimes had acquired both their languages in early childhood (e.g., Anooshian & Hertel, 1994; Caldwell-Harris et al., 2011; Harris et al., 2006). In these contexts, the L2 was the majority language in society and the L1 the minority language used at home, which can then be understood as heritage language (HL, Montrul, 2008). Others worked with populations who had learned an L2 as a foreign language in school (e.g., Iacozza et al., 2017; Thoma & Baum, 2019; Winskel, 2013).

Conceptually, language-dependent emotion differences could be confined to early automatic semantic processing that change or vanish in later and more conscious stages of emotion processing (Opitz & Degner, 2012; Thoma, 2021; Thoma & Baum, 2019; Winskel, 2013). Opitz and Degner (2012), for example, observed similar amplitudes in event-related potentials when bilinguals read L1 or L2 emotion words, but the reactions were delayed in L2. In this respect, emotions can generally be understood as an individual's experiential, physiological, and behavioral reactions to personally relevant stimuli (Mauss & Robinson, 2009), whereas feelings are (meta-)cognitive experiences of emotions (Damasio & Carvalho, 2013).

Table 1. Systematic review of multi-measure studies on bilinguals' language-dependent emotions.

Study	Bilingual sample	Main task	Finding about emotional reaction	Self-report (meta-cognitive) measure
Anooshian and Hertel (1994)	English-Spanish (18) and Spanish-English (18) in the United States	Word reading	Neurophysiological or cognitive measure	Verbal ratings L1 ~ L2
Harris et al. (2003)	Turkish-English (32) in the United States	Word reading	Word recall Emotion > neutral words but only in L1	Valence rating L1 ~ L2
Harris et al. (2006)	Spanish-English in the United States, early (31) vs late (21) English learners	Word reading	SCR L1 > L2	Valence rating L1 ~ L2
Caldwell-Harris and Aycıteğ-Dinn (2009)	Turkish-English (70) in Turkey	Phrase reading	SCR L1 ~ L2 L1 > L2	Verbal rating L1 > L2, interaction with word emotionality
Caldwell-Harris et al. (2011)	Mandarin-English (64) in the United States	Sentence listening (elicit autobiographical memory)	SCR L1 ~ L2 (some variation with proficiency)	Verbal rating L1 < L2 for reprimands (language × emotionality interaction)
Winkel (2013)	Thai-English (57) in Thailand	Word rating and emotional Stroop task	Stroop task L1 > L2	Verbal rating L1 ~ L2
Jonczyk et al. (2016)	Polish-English (19) in the United Kingdom (late immersed)	Sentence reading	ERP L1 > L2 for negative valence (language × emotionality interaction)	Valence rating L1 ~ L2
Iacozza et al. (2017)	Spanish-English (54) in Spain	Sentence reading	Pupillometry L1 > L2 for high emotionality (language × emotionality interaction)	Verbal rating L1 ~ L2
Jankowiak and Korpai (2018)	Polish-English (27) in Poland	Narrative text reading	SCR L1 > L2	Verbal rating L1 ~ L2
Thoma (2021) Exp. 2	German-English (130) in Germany	Sentence listening	Pupillometry L1 > L2 (neutral) L1 < L2 (rumination)	Verbal rating L1 ~ L2 (neutral) L1 > L2 (rumination)
(Thoma) et al. (2022)	German-English (66) in Germany	Sentence reading, listening to narrative	Pupillometry L1 > L2 for high arousal (language × emotionality interaction)	Verbal rating L1 ~ L2

Note. SCR: skin conductance resonance; ERP: event-related potentials; ">" "<"; significantly larger/smaller emotional reaction; "~": no significant difference.

There are at least three theoretical approaches explaining language-dependent emotions (see also Williams et al., 2020). First, *context theory* assumes a strong association between emotional learning and habitual language use contexts. As L2s are typically learned later and used in less emotional contexts, they will be associated with less intense emotionality than L1s (Ayçiçeği-Dinn & Caldwell-Harris, 2009; Brase & Mani, 2017; Iacozza et al., 2017). The approach emphasizes the joint age of acquisition (AoA) of emotion and language. It therefore predicts that L1 stays the most emotional language in a bilingual's life even if they start using L2 in similarly emotional contexts and become more proficient. Second, *processing theory* assumes that as L2 processing is less efficient and poses a higher cognitive load, it will hinder emotion processing, so that emotional reactions to/in L2 are less intense at a particular time (Opitz & Degner, 2012; Sianipar et al., 2015; Thoma & Baum, 2019). Accordingly, language-dependent emotion differences will diminish with increasingly balanced proficiency in both languages. Note that context and processing-based theories are interconnected because if an L2 is acquired earlier and used more frequently, it is more likely to coincide with emotional experiences, and its processing becomes more automatic (Pavlenko, 2012). Third, *cultural frame-switching theory* presumes that specific languages evoke socioculturally motivated patterns of behavior because they are associated with different sociocultural expectations (e.g., Panayiotou, 2004; Ross et al., 2002; Zhou et al., 2021). If these expectations involve affect, for example, in that Chinese-English bilinguals in the United States report more positive mood when asked in English than Chinese (Ross et al., 2002), the theory is also called *emotional frame switching* (Zhou et al., 2021). The approach resembles context theory in its emphasis on experience. However, even if context theory expects equivalence because bilinguals have acquired both languages simultaneously and use them in similar contexts comparably often, frame switching could accommodate language-dependent emotions provided a language evokes specific frames associated with the status of itself or the speech community (He, 2010). Such framing is already reflected in the ideologically biased but psycholinguistically vague term *native language* (Dewaele et al., 2021). Accordingly, language-induced emotion differences should be most pronounced at the level of conscious self-reported feelings than in automatic emotional reactions (Levenson et al., 2010), and if a bilingual's languages can trigger different frames—which has been shown for majority and HLs (e.g., Noriega & Blair, 2008).

Against this background, the present research investigated if differential language-dependent emotion effects inferred from an automatic measure of emotional responsiveness (here pupillometry) and self-consciously rated feelings (here arousal and valence icon scales) are methodological noise or indicative of different stages of bilingual emotion processing.

The present study

The aim of this study was to compare balanced simultaneous and unbalanced sequential bilinguals' automatic versus conscious emotional reactions to an emotional narrative told in different languages. We operationalized bilingualism in terms of AoA, language use frequency, and proficiency (Treffers-Daller, 2019). The balanced simultaneous bilinguals should be HL speakers. Due to migration, Turkish and Russian are the most frequent heritage (or minority) languages in Germany (Olfert & Schmitz, 2017). Communities with a Turkish HL background typically consider the HL their family language and have a positive attitude toward it (Bayram & Wright, 2017). Even though the situation is more variable within the Russian HL community, members who maintain Russian as a functional minority language also tend to value the language as part of their identity (for review, Olfert & Schmitz, 2017). The subsample of more unbalanced sequential bilinguals should use German as their first language (L1) and either English or French as a second language (L2)

predominantly in instructional, educational, or work contexts. When comparing it with German, we refer to the other language (HL or L2) as Lx.

We used pupil dilation during listening to the narrative as an automatic measure as it reflects unbiased reactions of the sympathetic nervous system to emotional stimuli (Bradley et al., 2008). Furthermore, it allows for more authentic emotion tasks than decision-based reaction time experiments and is less invasive than measuring skin or brain activity (Iacozza et al., 2017; Thoma & Baum, 2019).

Crossing these types of bilinguals and emotion measures allowed us to test conflicting predictions made by the three theories of language-dependent emotions. For HL speakers, context theory predicts that they show equal emotional reactions in German and Lx, provided the AoA and relevant language use contexts are similar. Processing theory also expects similar reactions from HL speakers, yet for the different reason that comparable language processing fluency enables comparable emotion processing. Frame-switching theory allows for language-dependent emotions even if AoA, emotion-relevant language use, and processing skills in German and Lx are comparable, but one language fits better with emotion-relevant expectations.

For unbalanced sequential L2 speakers, context theory predicts stronger automatic responses in German than in Lx because the earlier AoA of German in emotional contexts leads to stronger bodily “emotional resonance.” In principle, the automatic L1 emotion advantage should permeate to the subsequent conscious level of feelings. Note that authors who used context-based explanations and found incongruent automatic and conscious emotion differences have only provided methodological explanations, if any, for inconsistencies (e.g., Anooshian & Hertel, 1994; Caldwell-Harris & Ayçiçeği-Dinn, 2009; Iacozza et al., 2017). Processing theory assumes stronger automatic emotional reactions in L1 because semantic processing is less fluent in Lx. Yet, at the conscious rating level, language differences may disappear because language and emotion processing are sensitive to time and attention regulation (Opitz & Degner, 2012; Thoma, 2021). It is difficult to say if frame-switching induces automatic emotional reactions, but the theory certainly allows for language-dependent feelings because any language can trigger different attitudes and expectations (e.g., Edwards & Fuchs, 2018; Nederstigt & Hilberink-Schulpen, 2018). In sum, at least context and processing theory jointly lead to hypothesis H1:

H1: Balanced simultaneous bilinguals are less sensitive in their automatic pupillary reactions when exposed to a German versus Lx emotional stimulus than unbalanced sequential bilinguals are.

In contrast to context theory that does not expect variation between language-dependent automatic and self-conscious emotions, processing theory suggests that slower, more controlled, and deliberate processing reflected in self-reported feelings might modulate early automatic language-dependent emotion differences. Frame-switching theory predicts that language-induced differences are more distinct if languages are associated with distinct sociocultural values or attitudes. In conjunction, we, therefore, hypothesize:

H2: Balanced simultaneous bilinguals report larger differences in language-dependent feelings than unbalanced sequential bilinguals do.

Method

Participants. The study included 161 university students who received €10 as a thank you. They were from four different bilingual populations living in an urban area in Germany and proficient in

the majority language German. A group of Russian-German and Turkish-German bilinguals were from second- or third-generation immigrant families. Furthermore, a group of German-English and German-French bilinguals had initially learned English and French as foreign languages. We report the biographic and bilingualism statistics of the four subsamples at the beginning of the “Results” section.

Materials. We used a 10-second street traffic surveillance video as a baseline for pupil size under unemotional yet language-dependent processing load. Its audio track contained a traffic update stating there were no special incidents in the region without using emotion or emotion-laden words. The emotional narrative was presented as a high-arousal 58-second video called *Hochzeitstag* (“Wedding Day,” Minghella, 2006) previously used by Thoma and Baum (2019). It features an elderly lady in a cemetery who receives flowers from a delivery service her late husband had pre-ordered for their wedding day. The narrator reads an accompanying letter in the voice of the husband. The commercial won several awards for transporting love, dignity, and surprise. We used the German and English audio track spoken by the same balanced bilingual male actor with a mature, sonorant voice from Thoma and Baum (2019) and comparable speakers of French, Russian, and Turkish who read close translations of the narrative (see Supplement B for transcriptions). As the video was 14 years old and originally aired in Switzerland, it was unlikely to be known by participants. In qualitative pretests, representatives of all four bilingual subsamples evaluated the setting and portrayed interpersonal relationship as plausible and culturally appropriate. To have a control measure for individual differences in parasympathetic pupil reactions to light (Bradley et al., 2008), we estimated the mean perceived relative luminance per video second computed with the luminous efficiency function (Poynton, 2012, Eq. 24.1) integrating red, green, and blue tristimulus values of the individual still images.

Procedure. Participants were invited in an email newsletter and via personal contacts to a study about emotionality in international advertising. Upon arrival in the eye-tracking lab, they provided written informed consent and were then randomly assigned to the German or the matching Lx conditions. After assignment, all instructions and tasks were in the language of the condition or non-verbal. Participants were seated approximately 70 cm in front of a 24"-computer screen connected to a remote SMI RED 500 Hz eye-tracking system. Room lightning was artificial and constant at 590 lux. The experiment started with a 9-point eye-tracking calibration that was repeated until position accuracy was $\leq .5^\circ$ degrees of visual angle. To obtain a mood baseline and allow practicing response by eye fixation, participants indicated how they felt before the experiment on a simple Visual Analogue Mood Scale (VAMS; see van Rijsbergen et al., 2012 for review). They saw a 10-cm line with “happy” and “sad” poles, numbers 0 to 5 under the line, and five emojis on top and fixated the preferred emoji for 2 seconds to respond. Next, participants saw the baseline video followed by a 1-second fixation screen and the emotional target video (see Figure 2 for the timeline). Directly afterward, they read the instruction “How do you feel about this video?” in the language of the video and reported their response by fixating on one of the icons on 5-point SAM (Self-Assessment Manikin) arousal and valence scales (Lang, 1980). Participants then switched to another computer to answer a survey on their background data, AoA, language use, and proficiency self-assessment before they were paid and debriefed.

Pupil data preparation. A Python script pre-processed the pupillometry data according to conventional procedures (Kinner et al., 2017; Lemerrier et al., 2014; Thoma & Baum, 2019). The script deleted recordings during blinks and computed a mean diameter from the left and right pupil size for each measurement. Pupil dilation was defined as the difference between each participant’s

Table 2. Subsamples of bilinguals and their AoA of German and their other language (Lx).

Sample	N	Age		Female	AoA German (years)		AoA Lx (years)	
		M	SD	%	M	SD	M	SD
RUS-GER	40	22.98	3.37	65	1.37	1.75	0.85	2.08
TUR-GER	32	22.91	3.74	64	1.50	1.61	1.53	3.36
GER-ENG	58	21.17	2.23	64	0.00	0.00	8.79	2.61
GER-FRE	31	23.06	3.44	64	0.00	0.00	9.23	2.62

Note. AoA: age of acquisition.

pupil diameter during the target video and their mean pupil size observed during the baseline video, so that cognitive load of processing audio input in a particular language should be controlled. To balance for measurement noise, the script finally down-sampled the dilation values from 500 Hz to 1-second epochs. Data analysis was confined to a time of interest during the target video. It started 1 second after the narrator voice set in and stopped 3 seconds after the voice ended (see Figure 2) to include only epochs where language could have made a difference.

Results

Validation of sampling. Before testing the hypotheses, we validated if the four subsamples qualified as HL speakers and L2 speakers considering self-reported AoA, language use frequency in emotional contexts, and self-assessed language proficiency. Due to the large number of theoretically possible comparisons, we only report selected differences.

First, Table 2 shows that the Russian-German and Turkish-German speakers had acquired both of their languages in early childhood in family settings, and their AoA did not differ significantly between the languages, $F(1, 70)=0.37, p=.543, \eta^2=.01$, and the subsamples, $F(1, 70)=0.47, p=.494, \eta^2=.01$. The German-English and German-French participants had acquired German from birth in their families, while they had initially learned English or French as a foreign language in school. The AoA differences of about 9 years between their L1 and L2 were significant, $F(1, 88)=1046.40, p<.001, \eta^2=.92$, while the AoA of English and French was similar, $t(87)=-0.74, p=.459, d=-0.17$.

Second, we estimated language use frequency in emotional contexts for each language from the responses to two-item ascending 7-point rating scales. For each language, participants rated its use in contexts with family and friends (i.e., presumably emotional contexts) as well as in work and educational contexts. From these ratings, we computed a ratio score indexing the use of Lx in each context relative to the use of German (ratio=Lx/(Lx + German)). A ratio of 1 indicates exclusive use of Lx in a context. Figure 1 plots the mean ratios for each sample of bilinguals showing, for example, that German was clearly the dominant language in work and education settings for all groups. Importantly, a repeated measures analysis of variance (ANOVA) confirmed that whereas the use of German or Lx differed significantly between family-friends and work-education contexts for all bilingual samples, $F(1, 157)=44.57, p<.001, \eta^2=.22$, the pattern was reverse for HL and L2 speakers, $F(1, 157)=89.76, p<.001, \eta^2=.63$. In fact, both HL samples used Lx more often in emotional family-friends contexts than for work and education (compare the first and second bars, respectively, in Figure 1), Russian: $t(39)=13.66, p<.001, d=0.14$; Turkish: $t(31)=7.15, p<.001, d=0.14$. Both L2 samples, in contrast, used Lx less frequently with family and friends

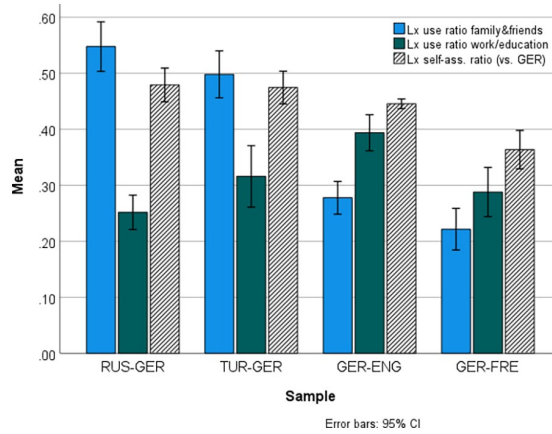


Figure 1. Means and confidence intervals for the language use frequency and self-assessment ratio scores across the four subsamples and their languages. A ratio of 0.50 indicates balance between German and Lx.

than for work and education, English: $t(57)=-5.98$, $p<.001$, $d=0.15$; French: $t(30)=-3.43$, $p=.002$, $d=0.11$.

Third, participants self-assessed their speaking, reading, writing, listening, and grammar skills on 7-point scales in both of their languages. The ratings were first transformed into a proportionate mean proficiency score ranging from 0 (*lowest*) to 1 (*highest*) and then into an Lx self-assessment ratio score relative to German in analogy to the use ratio reported above. If the ratio is 0.5, this indicates balanced proficiency in Lx and German. As the striped bars in Figure 1 suggest, HL speakers' German and Lx proficiency were similar, whereas L2 English and French skills were lower, compared to L1 German. A repeated measures ANOVA comparing Lx versus German self-assessments confirmed a main effect of language, $F(1, 157)=104.74$, $p<.001$, $\eta^2=.40$, and an interaction with sample, $F(3, 157)=17.34$, $p<.001$, $\eta^2=.25$. In paired comparisons, Russian, $t(39)=1.20$, $p=.237$, $d=0.19$, and Turkish, $t(31)=1.79$, $p=.082$, $d=0.32$, speakers rated their Lx proficiency similar to German. In contrast, the German-English, $t(51)=13.48$, $p<.001$, $d=1.77$, and German-French, $t(51)=9.22$, $p<.001$, $d=1.67$, sample evaluated their proficiency in Lx substantially lower than in German.

Based on the AoA, language use, and proficiency data, we concluded that the Russian-German and Turkish-German HL subsamples were balanced simultaneous bilinguals. As expected, the German-English and German-French samples were sequential unbalanced L2 speakers.

Pupil dilation. As the line plot in Figure 2 illustrates, pupil dilation tends to evolve over time, and it is generally highly susceptible to individual differences (Lemercier et al., 2014). Therefore, we analyzed the pupil data with linear mixed-effects regression models using the *lmer* function from the *lme4* package (Bates et al., 2021) and *lmerTest* (Kuznetsova et al., 2017) to estimate p values in RStudio (RStudio Team, 2021). The categorical fixed factors were deviation-coded with bilingual status (HL=-.5 vs L2=.5) and language of the video (LoV; German=-.5 vs Lx=.5). The continuous co-variates mood baseline, epoch, and luminance were centered to achieve comparable regression estimates. Pre-experimental mood was similar in the four comparison groups (see Table 4). Interactions beyond the two categorical predictors did not improve model fit (backward fitting

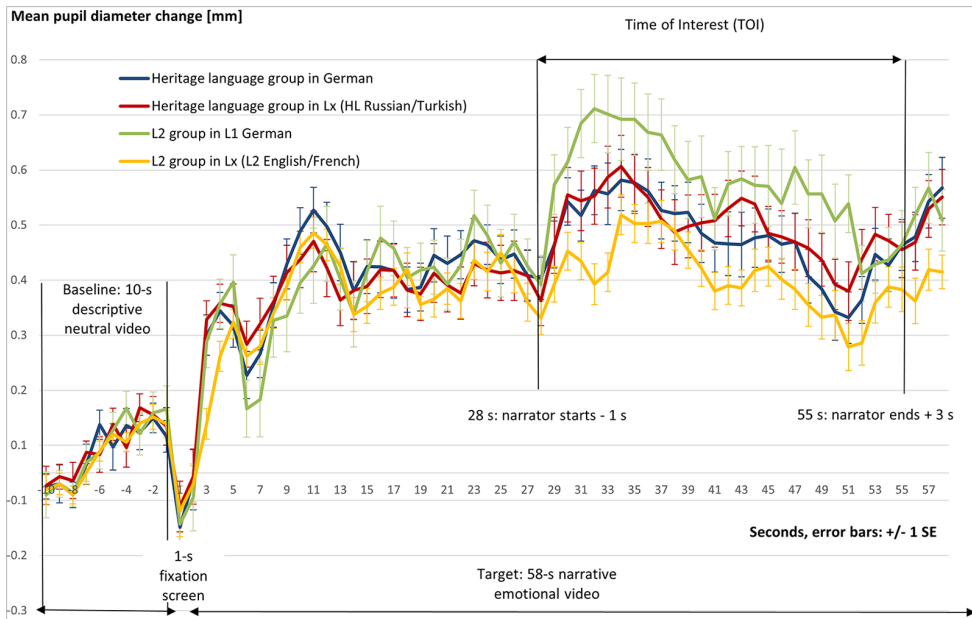


Figure 2. Pupil diameter change as a function of time (epoch), bilingual status, and language of the video.

via log-likelihood comparisons; Barr et al., 2013). All models included by-participant random intercepts and slopes for bilingual status and LoV, and a random intercept for epoch.

The model predicting pupil dilation in Table 3 revealed only a significant main effect of epoch. Pupil dilation showed the typical, decreasing pattern after an initial peak when the narration set in across all conditions (see Figure 2). Notably, bilingual status and LoV interacted significantly as visualized in Panel (a) of Figure 3. Separate models fitted to the subsample data split by bilingual status showed that while heritage speakers' pupil responses were very similar in German and Lx, $b=0.020$, $SE=0.063$, $t=0.32$, $p=.751$, L2 speakers showed significantly stronger pupil dilation in German than in Lx, $b=-0.174$, $SE=0.058$, $t=-3.02$, $p=.003$.

Self-reported arousal and valence. Table 4 summarizes the descriptive values for SAM arousal and valence reported by the four comparison groups. Note that the valence scale was bi-polar ($-2=sad$, $2=happy$) and therefore recoded, so that larger numbers symbolize stronger negative or positive valence. We used ordinal logistic regression models to predict arousal and valence ratings (polr function from MASS package, version 7.3-54; Venables & Ripley, 2007). Predictors were deviation-coded bilingualism status, LoV, and their interaction as well as the centered mood score.

The model predicting arousal (Table 5) found a significant main effect of bilingual status because the L2 group stated overall higher arousal than the HL group. The main effect of LoV was also significant with higher ratings in Lx than German. These main effects were qualified by a significant interaction (see Panel (b) of Figure 3), such that HL speakers experienced language-dependent arousal differences, $b=1.735$, $SE=0.475$, $t=3.65$, $p<.001$, whereas L2 speakers did not, $b=0.049$, $SE=0.435$, $t=.11$, $p=.910$. In the valence model (Table 5), LoV showed a significant main effect with Lx eliciting stronger positive and negative feelings. The effect of bilingual status and the two-way interaction did not reach significance even though the valence advantage of

Table 3. Linear mixed-effects regression model predicting pupil dilation.

Dependent variable	Pupil dilation (mm)			
	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Fixed effects				
(Intercept)	0.441	0.022	20.12	0.000
Bilingual status	0.005	0.042	0.12	0.906
Language of video (LoV)	-0.077	0.043	-1.82	0.071
Mood baseline (VAMS)	0.012	0.020	0.57	0.568
Epoch	-0.042	0.009	-4.84	0.000
Luminance	-0.012	0.009	-1.41	0.170
Bilingual × LoV	-0.198	0.085	-2.33	0.021

Note. Model: $\text{lmer}(\text{Pupil dilation} \sim 1 + \text{Bilingual status} * \text{LoV} + \text{scale}(\text{VAMS}) + \text{scale}(\text{Epoch}) + \text{scale}(\text{Luminance}) + (1 | \text{ID: Bilingual status: LoV}) + (1 | \text{Epoch}))$. VAMS: Visual Analogue Mood Scale.

Table 4. Means and standard deviations for mood, arousal, and valence.

Bilingual status	Language of video	VAMS mood (0–5)		SAM arousal (1–5)		SAM valence (0–3)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
HL	German	2.61	0.93	2.64	0.83	0.83	0.78
	Lx	2.72	0.91	3.58	1.02	1.42	0.73
L2	German	2.48	0.89	3.44	0.80	0.85	0.60
	Lx	2.66	0.79	3.45	0.76	1.08	0.73

Note. VAMS: Visual Analogue Mood Scale; SAM: Self-Assessment Manikin (Lang, 1980); *SD*: standard deviation; HL: heritage language speakers; L2: second language speakers.

Lx relative to German was significant in the HL group, $b=1.465$, $SE=0.471$, $t=3.11$, $p=.001$, but weak in the L2 group, $b=0.751$, $SE=0.444$, $t=1.69$, $p=.091$.

Discussion

We examined if balanced simultaneous bilinguals and unbalanced sequential bilinguals respond differently to an emotional narrative in their languages in automatic and self-conscious measures of emotion. Based on self-reported AoA, language use, and proficiency, we defined two samples of Russian and Turkish HL speakers as balanced simultaneous bilinguals because they had acquired German and the Lx in early childhood, used both languages comparably often in presumably emotional contexts with family and friends, and had similar overall proficiencies. Two samples of German-English and German-French speakers qualified as unbalanced sequential bilinguals because they had acquired Lx English or French as a foreign language substantially later than their L1 German, reported to use Lx predominantly in educational and work contexts, and rated their Lx proficiency lower.

The two types of bilinguals showed a reversed pattern of language-dependent emotional reactions measured in terms of automatic pupillary responsiveness, compared to self-reported arousal and valence ratings. Heritage speakers' pupil dilation was similar in both languages, while they rated the German narrative of the same story less emotional than the Lx one. The main and interaction effects of bilingual status were significant for arousal and indicated a similar trend for valence ratings. Since language-dependent emotion differences seem to increase with task emotionality (Thoma, 2021) and the story in the video is predominantly heart-warming and moving rather than pleasant or unpleasant, this discrepancy between arousal and valence could be a task effect.

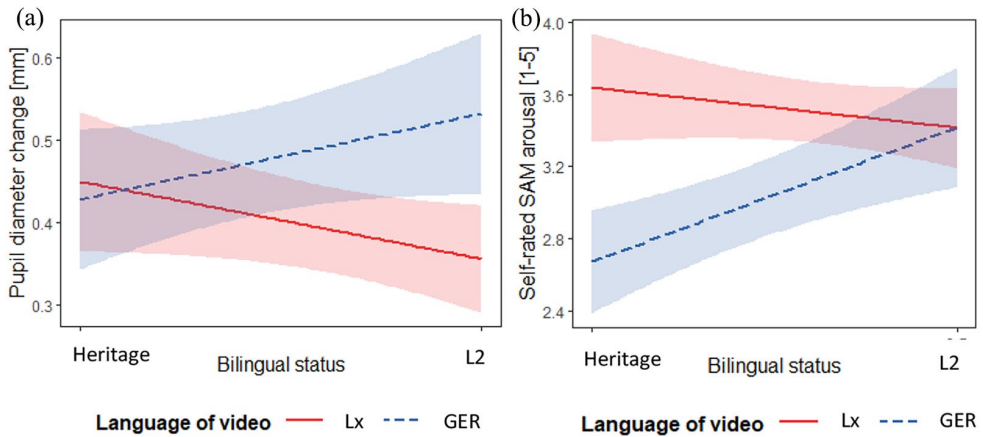


Figure 3. Co-variate adjusted interaction plots for bilingual status \times language of the video with 95% CIs for (a) pupil dilation and (b) self-rated arousal. Labels: Heritage: group of Russian and Turkish heritage speakers; L2: group of English and French second language speakers; Lx: language other than German (Heritage or L2); GER: German.

Table 5. Linear regression models predicting self-reported arousal and valence.

Dependent variable	SAM arousal				SAM valence			
	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Bilingual status	0.706	0.316	2.24	0.025	-0.458	0.313	-1.46	0.144
Language of video (LoV)	1.004	0.318	3.16	0.002	1.171	0.325	3.61	0.000
Mood baseline (VAMS)	0.126	0.150	0.84	0.398	-0.274	0.150	-1.82	0.069
Bilingual status \times LoV	-1.956	0.633	-3.09	0.002	-1.086	0.626	-1.73	0.083

Note. VAMS: Visual Analogue Mood Scale; SAM: Self-Assessment Manikin (Lang, 1980).

Importantly, L2 speakers showed weaker automatic reactions in Lx than German, yet they rated the narrative similarly emotional in both languages.

We interpret these results in the light of two hypotheses. First, prior work designed against the background of context and processing theories of bilingual emotions has rarely sampled balanced simultaneous bilinguals because they should not show any interesting differences. The present results confirmed this with the data obtained from the automatic measure. Also consistent with prior studies (e.g., Iacozza et al., 2017; Thoma & Baum, 2019), our unbalanced sequential bilinguals' pupil reactions were stronger in German than in Lx suggesting stronger automatic emotional reactions in German. Our experimental paradigm integrated both populations. It revealed a significant interaction between the listeners' bilingual status and the language of the emotional narrative. This confirmed our H1, in that balanced simultaneous (HL) bilinguals' automatic pupillary reactions were less sensitive in response to a German versus Lx emotional stimulus, compared to unbalanced sequential (L2) bilinguals.

In parallel to most multi-measure studies reviewed in Table 1, the language-induced difference at the automatic level did not permeate to self-reported arousal and valence. This is difficult to explain for context theory because the later AoA and less emotional use contexts of Lx should lead

to persistently weaker emotions in Lx, that is, L2. Processing theory, in contrast, can accommodate this disparity. Here, less efficient L2 language processing interferes with early automatic emotion processing, which leads to disparate emotional reactivity in L1 and L2. Self-reported feelings, however, also represent the outcome of later, more deliberate stages of emotion processing, where L2 processing catches up (Opitz & Degner, 2012) and different emotion regulation mechanisms are active (Thoma, 2021), so that the initial language-dependent disparity cannot always be observed. To further validate these findings, future work could include unbalanced simultaneous and balanced sequential bilinguals.

Second, context and processing theory would not have predicted that balanced simultaneous bilinguals rate the narrative less (or more) emotionally intense in their majority language German than in their HL Russian or Turkish. Within a frame-switching account, an HL can trigger sociocultural expectations, so that an HL narrative *feels* more emotional (Noriega & Blair, 2008; Zhou et al., 2021). Several multi-measure studies (Table 1) did not find a main effect of language at the level of reported feelings. Therefore, we hypothesized (H2) that balanced simultaneous bilinguals report larger differences between feelings induced by German versus HL than unbalanced sequential bilinguals between German versus L2. A significant interaction in the arousal model and a similar trend for valence confirmed H2. HL speakers rated the narrative more emotional in their HL presumably because it triggered the frame of a *native* language or *mother tongue* (Dewaele et al., 2021). The HL frame created a more appropriate “emotional fit” (Zhou et al., 2021) with the situation presented in the emotional narrative video than German although experience and learning have trained our balanced simultaneous bilinguals to process emotional stimuli equally well in both languages. Our results do not warrant claims about frames that are specific to the Russian or Turkish heritage culture, but the parallel effects in both HL groups suggest that the sociocultural function of HL in general (He, 2010) influences self-reported feelings. Further research could substantiate this interpretation by eliciting and evaluating language-dependent frames.

In fact, positive attitudes toward L2 English and French (e.g., Edwards & Fuchs, 2018) and expectations about the language choice in advertisements (Nederstigt & Hilberink-Schulpen, 2018) could also explain why the reduced automatic emotional reactivity in L2 did not surface in self-conscious feelings because positive attitudes offset the disparity at the automatic level. However, there are at least five other explanations for language-independent feelings in the L2 groups. Initial differences (1) may disappear after long enough processing (Opitz & Degner, 2012), or (2) they could be offset by emotion regulation (Thoma, 2021). The L2 (3) may also trigger different processing modes (de Langhe et al., 2011) or (4) decision modes (Keysar et al., 2012), and (5) reduced familiarity of L2 could stipulate a novelty effect that made the L2 narrative more attractive (Ayçiçeği-Dinn & Caldwell-Harris, 2009). The trend toward higher perceived valence in L2 may indicate interest and attitude effects, but our research is limited in that we did not collect the data to test those explanations, for example, on language attitudes, processing mode, or interest in the narrative video. While this could be addressed in future research, the limitation that we did not use the same language pairs in the HL and L2 groups is more difficult to overcome because Turkish is not taught as a foreign language in regular German schools and French heritage speakers are rare in Germany, for example. It may be promising, however, to work with multiple emotion stimuli that also vary socioculturally to disentangle effects of language and culture.

In combination, the current findings support that automatic neurophysiological and conscious meta-cognitive measures of emotion tap into different stages or dimensions of emotion processing or, in other words, language-induced differences between intuitive emotional reactions and feelings. Theories of bilingual emotions need further development to explain these incongruent emotional reactions consistently and explicitly. Methodologically, the findings advocate for multi-measure approaches to enhance the validity of future research on bilingual emotions.

Acknowledgements

The author thanks Amanda Baum, Nina Harlacher, Seda Ince, Sarah Siebenmorgen, and Christina Weninger for their help with designing materials, recruiting and testing participants.

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.

Data availability statement

The data sets of this experiment are available at https://osf.io/ax9b5/?view_only=ec5cc00a67a8428aa14be3b01d08e76b.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Dieter Thoma  <https://orcid.org/0000-0001-6984-623X>

Supplemental material

Supplemental material for this article is available online.

References

- Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition and Emotion*, 8(6), 503–514. <https://doi.org/10.1080/02699939408408956>
- Ayçiçeği-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, 12(3), 291–303. <https://doi.org/10.1017/S1366728909990125>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Bojesen Christensen, R. H., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., & Fox, J. (2021). *Package lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-27.1) [Computer software]. <https://cran.r-project.org/package=lme4>
- Bayram, F., & Wright, C. (2017). Turkish heritage language acquisition and maintenance in Germany. In P. P. Trifonas & T. Aravossitas (Eds.), *Handbook of research and practice in heritage language education* (pp. 481–502). Springer. https://doi.org/10.1007/978-3-319-44694-3_49
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607. <https://doi.org/10.1111/j.1469-8986.2008.00654.x>
- Brase, J., & Mani, N. (2017). Effects of learning context on the acquisition and processing of emotional words in bilinguals. *Emotion*, 17(4), 628–639. <https://doi.org/10.1037/emo0000263>
- 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>
- Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, 14(2), 143–152. <https://doi.org/10.1038/nrn3403>

- de Langhe, B., Puntoni, S., Fernandes, D., & van Osselaer, S. M. J. (2011). The anchor contraction effect in international marketing research. *Journal of Marketing Research*, 48(2), 366–380. <https://doi.org/10.1509/jmkr.48.2.366>
- Dewaele, J.-M. (2016). Thirty shades of offensiveness: L1 and LX English users' understanding, perception and self-reported use of negative emotion-laden words. *Journal of Pragmatics*, 94, 112–127. <https://doi.org/10.1016/j.pragma.2016.01.009>
- Dewaele, J.-M., Bak, T. H., & Ortega, L. (2021). Why the mythical “native speaker” has mud on its face. In N. Slavkov, S. Melo-Pfeifer, & N. Kerschhofer-Puhalo (Eds.), *The changing face of the “native speaker”* (pp. 25–46). De Gruyter. <https://doi.org/10.1515/9781501512353-002>
- Edwards, A., & Fuchs, R. (2018). A cluster analysis of attitudes to English in Germany and the Netherlands. *World Englishes*, 37(4), 653–667. <https://doi.org/10.1111/weng.12348>
- Garrido, M. V., & Prada, M. (2021). Comparing the valence, emotionality and subjective familiarity of words in a first and a second language. *International Journal of Bilingual Education and Bilingualism*, 24(2), 275–291. <https://doi.org/10.1080/13670050.2018.1456514>
- Harris, C., Ayçiçeği-Dinn, 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>
- Harris, C., Gleason, J. B., & Ayçiçeği-Dinn, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual education and bilingualism. Bilingual minds: Emotional experience, expression and representation* (pp. 257–283). Multilingual Matters.
- He, A. W. (2010). The heart of heritage: Sociocultural dimensions of heritage language learning. *Annual Review of Applied Linguistics*, 30, 66–82. <https://doi.org/10.1017/S0267190510000073>
- Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLOS ONE*, 12(10), Article e0186027. <https://doi.org/10.1371/journal.pone.0186027>
- Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472), 1517–1522. <https://doi.org/10.1126/science.aaw8160>
- Jankowiak, K., & Korpil, P. (2018). On modality effects in bilingual emotional language processing: Evidence from galvanic skin response. *Journal of Psycholinguistic Research*, 47(3), 663–677. <https://doi.org/10.1007/s10936-017-9552-5>
- Jończyk, R., Boutonnet, B., Musiał, K., Hoemann, K., & Thierry, G. (2016). The bilingual brain turns a blind eye to negative statements in the second language. *Cognitive, Affective & Behavioral Neuroscience*, 16(3), 527–540. <https://doi.org/10.3758/s13415-016-0411-x>
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668. <https://doi.org/10.1177/0956797611432178>
- Kinner, V. L., Kuchinke, L., Dierolf, A. M., Merz, C. J., Otto, T., & Wolf, O. T. (2017). What our eyes tell us about feelings: Tracking pupillary responses during emotion regulation processes. *Psychophysiology*, 54(4), 508–518. <https://doi.org/10.1111/psyp.12816>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Ablex Publishing Corporation.
- Lemercier, A., Guillot, G., Courcoux, P., Garrel, C., Baccino, T., & Schlich, P. (2014). Pupillometry of taste: Methodological guide—from acquisition to data processing—and toolbox for MATLAB. *Tutorials in Quantitative Methods for Psychology*, 10(2), 179–195. <https://www.tqmp.org/RegularArticles/vol10-2/p179/p179.pdf>
- Levenson, R. W., Soto, J., & Pole, N. (2010). Emotion, biology, and culture. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 780–797). Guilford Press.

- 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>
- Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, 5(2), 175–190. <https://doi.org/10.1037/1528-3542.5.2.175>
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, 23(2), 209–237. <https://doi.org/10.1080/02699930802204677>
- Minghella, A. (Director). (2006). *Hochzeitstag*. [Wedding Day] Fleurop-Interflora AG (Schweiz). <https://www.fleurop.ch/de/s/werbung>
- Montrul, S. A. (2008). *Incomplete acquisition in bilingualism* (Vol. 39). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.39>
- Nederstigt, U., & Hilberink-Schulpen, B. (2018). Advertising in a foreign language or the consumers' native language? *Journal of International Consumer Marketing*, 30(1), 2–13. <https://doi.org/10.1080/08961530.2017.1363008>
- Noriega, J., & Blair, E. (2008). Advertising to bilinguals: Does the language of advertising influence the nature of thoughts? *Journal of Marketing*, 75(5), 69–83. <https://doi.org/10.1509/jmkg.72.5.069>
- Olfert, H., & Schmitz, A. (2017). Heritage language education in Germany: A focus on Turkish and Russian from primary to higher education. In P. P. Trifonas & T. Aravossitas (Eds.), *Handbook of research and practice in heritage language education* (pp. 397–415). Springer International Publishing. https://doi.org/10.1007/978-3-319-44694-3_25
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967. <https://doi.org/10.1016/j.neuropsychologia.2012.04.021>
- Panayiotou, A. (2004). Switching codes, switching code: Bilinguals' emotional responses in English and Greek. *Journal of Multilingual and Multicultural Development*, 25(2–3), 124–139. <https://doi.org/10.1080/01434630408666525>
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405–428. <https://doi.org/10.1080/00207594.2012.743665>
- Poynton, C. (2012). *The Morgan Kaufmann Series in Computer Graphics. Digital video and HD algorithms and interfaces* (2nd ed.). Elsevier Science.
- Puntoni, S., de Langhe, B., & van Osselaer, S. M. J. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research*, 35(6), 1012–1025. <https://doi.org/10.1086/595022>
- Ross, M., Xun, W. Q. E., & Wilson, A. E. (2002). Language and the bicultural self. *Personality and Social Psychology Bulletin*, 28(8), 1040–1050. <https://doi.org/10.1177/01461672022811003>
- RStudio Team. (2021). *RStudio: Integrated development environment for R* (Version 2021.9.1.372) [Computer software]. RStudio, PBC. <http://www.rstudio.com/>
- Sianipar, A., Middelburg, R., & Dijkstra, T. (2015). When feelings arise with meanings: How emotion and meaning of a native language affect second language processing in adult learners. *PLOS ONE*, 10(12), Article 0144576. <https://doi.org/10.1371/journal.pone.0144576>
- Thoma, D. (2021). Emotion regulation by attentional deployment moderates bilinguals' language-dependent emotion differences. *Cognition and Emotion*, 35(6), 1121–1135. <https://doi.org/10.1080/02699931.2021.1929853>
- Thoma, D., & Baum, A. (2019). Reduced language processing automaticity induces weaker emotions in bilinguals regardless of learning context. *Emotion*, 19(6), 1023–1034. <https://doi.org/10.1037/emo0000502>
- Thoma, D., Hüsam, J., & Wielscher, K. (2022). Introducing grip force as a nonverbal measure of bilingual feelings. *Bilingualism: Language and Cognition*, 26, 125–137. <https://doi.org/10.1017/S1366728922000396>
- Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, 5(1), 375–393. <https://doi.org/10.1146/annurev-linguistics-011817-045554>
- van Rijsbergen, G. D., Bockting, C. L. H., Berking, M., Koeter, M. W. J., & Schene, A. H. (2012). Can a one-item mood scale do the trick? Predicting relapse over 5.5-years in recurrent depression. *PLOS ONE*, 7(10), Article e46796. <https://doi.org/10.1371/journal.pone.0046796>
- Venables, W. N., & Ripley, B. D. (2007). *Statistics and computing. Modern applied statistics with S* (4th ed.). Springer.

- Williams, A., Srinivasan, M., Liu, C., Lee, P., & Zhou, Q. (2020). Why do bilinguals code-switch when emotional? Insights from immigrant parent-child interactions. *Emotion, 20*(5), 830–841. <https://doi.org/10.1037/emo0000568>
- 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>
- Zhou, C., Dewaele, J.-M., Ochs, C. M., & de Leersnyder, J. (2021). The role of language and cultural engagement in emotional fit with culture: An experiment comparing Chinese-English bilinguals to British and Chinese monolinguals. *Affective Science, 2*(2), 128–141. <https://doi.org/10.1007/s42761-021-00037-x>

Author biography

Dieter Thoma is Senior Researcher at the Chair of Psycholinguistics in the Department of English at the University of Mannheim. He is interested in language acquisition and (bilingual) language processing. He has been studying how (second) language processing interacts with cognitive, emotional, and social processes in decision making and consumer behavior. He uses different experimental methods including reaction-time tasks, eye tracking and pupillometry as well as psychometric studies.