Title
When Emotion Slows You Down: Valence-Arousal Conflicts Across L1 and L2

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Stankovic, Michelle
Hamamura, Takeshi
Biedermann, Britta

Publication Date
2026

When Emotion Slows You Down: Valence-Arousal Conflicts Across L1 and L2

Michelle Stankovic (michelle.stankovic@upf.edu)[1], **Takeshi Hamamura**[2] **& Britta Biedermann**[3]
[1]Department of Language & Translation Sciences, Pompeu Fabra University
[2]School of Population Health, Curtin University, Bentley, Australia
[3]School of Allied Health, Curtin University, Bentley, Australia

Abstract

Across two experiments, we tested whether valence and arousal interact during speeded affective judgements, and whether this interaction differed between pictures and words, or between L1 and L2 words. German–English sequential bilinguals immersed in an L1 environment completed an affective judgement task. Experiment 1 focused on valence judgements (e.g., positive vs. negative) whereas Experiment 2 focused on arousal judgements (e.g., arousing vs. calming). In Experiment 1, a Valence–Arousal interaction pattern was present only within picture stimuli. In Experiment 2, this pattern was evident for all stimulus types showing that task demands shaped when the interaction was observable in words. Across both experiments, the interaction did not differ between L1 and L2 word blocks, suggesting that both languages were processed similarly. These findings suggest that speeded affective appraisals were not strongly language specific in highly proficient bilinguals. L2-only analyses showed that bilingual experience variables moderated response times.

Keywords: bilingual; second-language processing; emotion; valence; arousal; foreign language effect

Introduction

Cognitive processing is often evaluated in terms of speed and accuracy, but response times can also be informative about underlying processing computations. Within affective word processing, systematic inefficiency may reflect the integration of multiple evaluative cues rather than simple performance costs. The present paper examines how two foundational dimensions of affective meaning (valence, indicating whether a stimulus is positive or negative, and arousal, indicating the intensity of the stimulus) impact processing across representational formats (pictures vs. words) and languages (L1 vs. L2) in bilingual readers.

Affective Word Processing

Emotions help humans to effectively process information, and affective stimuli are well known to be processed differently than neutral stimuli (e.g., Bradley et al., 2001; Zajonc, 2001). A vast literature suggests a processing advantage for affective content, showing it is processed more efficiently compared to neutral stimuli. A large portion of this work has focused on pictorial stimuli (e.g., Feng et al., 2012), likely because pictorial stimuli have a direct sensory impact (e.g., not relying on imagination), and therefore larger capacity to induce strong physiological activity (Lang et al., 1997), thus producing larger and more consistent experimental effects.

However, these effects are harder to characterise for language. Words require lexical access and vary on properties that strongly affect response time, such as frequency and length (Brysbaert et al., 2018). This makes it easier for lexical variance to mask or reverse small affective effects in word recognition tasks.

Even so, facilitation effects have been reported for affective words in lexical decision and related paradigms (Kousta et al., 2009; Conrad et al., 2011; Ponari et al., 2015; see Hinojosa et al., 2020 for a review). However, the boundary conditions of this facilitation remain unclear, as findings are mixed across tasks and stimulus sets (Hinojosa et al., 2020). Some work reports early facilitative effects for negative words (e.g., Hofmann et al., 2009), and some for positive words (e.g., Barriga-Paulino et al., 2022). Other work emphasises that emotion effects depend on lexical factors such as frequency and concreteness and can be reduced or absent under some conditions (Ferré et al., 2018; Méndez-Bértolo et al., 2011).

This picture becomes more complex when assessing bilingual speakers. Despite research efforts, there is still no consensus on how words are assigned to emotional experiences and events across either their first or second language or how these lexicalisations become automatic (see Hinojosa et al., 2020 for a review). Bilingual word recognition differs from monolingual recognition because language experience is distributed across two systems and varies widely between individuals (Grosjean, 2008; De Bruin, 2019; Dijkstra et al., 2019). As a result, affective effects may interact with proficiency, age of acquisition, and language use, adding heterogeneity on top of the lexical variability already present in word tasks (De Bruin, 2019; Kaushanskaya et al., 2020).

Therefore, whether bilinguals process emotion similarly in their first language (L1) and second language (L2) remains unclear. This question is central to accounts of language-dependent emotion, including the foreign language effect, which proposes that using an L2 can attenuate emotional resonance and shift judgements (Caldwell-Harris, 2014; Pavlenko, 2012).

Recent work suggests that these affective dimensions do not act independently, rather interact to influence processing. Valence–Arousal Conflict Theory predicts faster responses for high-arousal negative stimuli and for low-arousal positive stimuli, and slower responses for conflicting combinations (high-arousal positive; low-arousal negative) (Robinson et al., 2004; Purkis et al., 2009; Citron et al., 2014a; Citron et

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

al., 2014b; Wang et al., 2018). This interaction has been reported for pictures and for words in monolingual samples, and it has also been linked to early neural differences during stimulus processing (Citron et al., 2013; Citron et al., 2014a).

The Present Studies

The present study uses a within-subject design to test how valence and arousal jointly shape speeded judgements across stimulus types (pictures vs. words) and languages (L1 vs. L2) in highly proficient bilingual participants. We utilised a tightly matched list of word and picture stimuli categorised across four distinct levels of valence (e.g., positive or negative) and arousal (e.g., high or low), based on norms from established corpora (e.g., Bradley & Lang, 1999; Bradley et al., 2001; Marchewka et al., 2014; Schmidtke et al., 2014).

The central question was whether the Valence–Arousal Conflict pattern is stable across stimulus type (pictures vs. words) and language nativeness (L1 vs L2), where the conflict refers to affective combinations that are less motivationally congruent: high-arousal positive and low-arousal negative stimuli. Our experiments test whether the Valence-Arousal Conflict Pattern depends on task demands and stimulus type, and whether it differs between L1 and L2 word blocks. Experiment 1 required a valence judgement, and Experiment 2 required an arousal judgement. Comparing these tasks examined whether the valence–arousal conflict reflects a stable property of affective processing, or whether it depends on the dimension participants must explicitly classify.

Methodology

Participants

Experiment 1 One hundred and sixty-eight German–English sequential bilinguals were recruited from Germany, either through Prolific or via convenience sampling at German universities. Ten participants were removed because German was not their L1, leaving 158 healthy adult participants for analysis. Participant characteristics are reported in Table 1.

Experiment 2 One hundred and thirty-nine German–English sequential bilinguals were recruited from Germany, either through Prolific or via convenience sampling at German universities. Six participants were removed because German was not their L1, leaving 133 healthy adult participants for analysis (see Table 1). The two samples were descriptively similar on age, education, L2 age of acquisition, LexTALE score, L2 reading proficiency, and daily L2 use, supporting comparison between the valence-judgement and arousal-judgement experiments.

Table 1: Participant characteristics

	Experiment 1 *Valence task*	Experiment 2 *Arousal task*
n	*158*	*133*
Age	31.25 (11.68)	32.13 (10.65)
Education	5.31 (1.39)	5.23 (1.46)
Lextale Score %	85.37 (11.68)	85.19 (11.72)
L2 Reading Proficiency	9.45 (1.17)	9.54 (0.99)
L2 Age of Acquisition	8.94 (2.65)	9.10 (3.05)
L2 Daily Use %	33.49 (21.26)	34.08 (22.46)
L2 family immersion	4.48 (8.48)	2.42 (7.53)
L1 Reading Proficiency	10.71 (0.57)	10.72 (0.56)
L1 Age of Acquisition	1.15 (1.40)	1.30 (1.57)
L1 Daily Use %	61.80 (22.67)	59.99 (23.52)
L1 family immersion	27.50 (9.56)	28.50 (9.93)

Note. Mean (SD) values presented. Education values represent range from (1) less than high school to (8) Ph.D./M.D./J.D. Reading proficiency was measured on an 11-point scale. Age of acquisition and L2 family immersion measured in years.

Materials

The same materials were used across both Experiments.

Word Stimuli Words were selected based on normative valence and arousal ratings from the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) and the Affective Norms for German Sentient Terms (ANGST; Schmidtke et al., 2014). The final set consisted of 40 translation-equivalent German–English word pairs, with ten items per Valence × Arousal cell (positive/negative × high/low arousal). The German and English words were matched pairwise on valence and arousal, as well as word length (number of letters) and word frequency. Word frequency was derived from SUBTLEX-US and SUBTLEX-DE words-per-million estimates (Brysbaert et al., 2011; Brysbaert et al., 2012). Because the English and German norms were reported on different scales, ratings for each measure were standardized as z-scores within each language prior to analysis.

The intended categories separated in the expected directions: positive stimuli had higher valence ratings than negative stimuli, and high-arousal stimuli had higher arousal ratings than low-arousal stimuli. A detailed description of the word selection and matching procedure is provided in Stankovic et al. (2024), including all means and SDs, for the relevant affective and psycholinguistic ratings described above. High arousal negative words included assault, blackmail, humiliate, hostage, cockroach, fear, destroy, annoy, scream, spider. Low arousal negative words included deformed, discouraged, urine, weary, muddy, feeble, dark, rigid, listless, aloof. High arousal positive words included aroused, lively, money, radiant, cheer, joy, admired, outstanding passion, decorate. Low arousal positive words included doll, hairpin, street, elevator, limber, custom, window, tree, nourish, tidy.

Picture Stimuli Forty-eight pictures were selected from the Nencki Affective Picture System (NAPS; Marchewka et al., 2014) based on normative valence and arousal ratings. Pictures were sampled to match the same Valence × Arousal structure as the word set, with twelve pictures per cell. Four example stimuli are presented in Figure 1. Further details about the picture ratings are available on OSF: https://doi.org/10.17605/OSF.IO/9JCFV

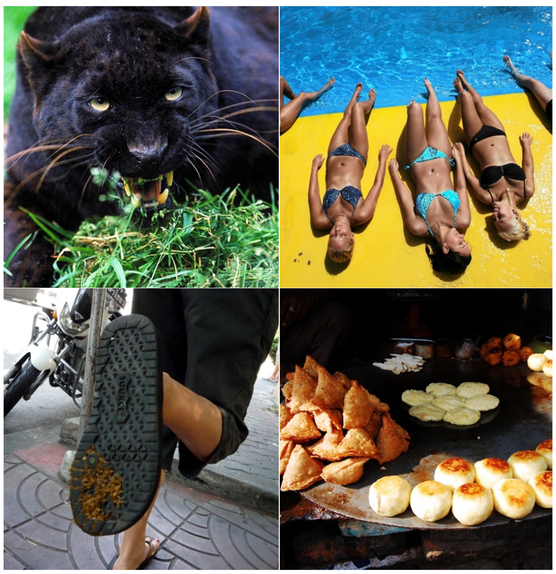

Figure 1: Example picture stimuli. Left panels show negative stimuli and right panels show positive stimuli; top panels show high-arousal stimuli and bottom panels show low-arousal stimuli.

Measures

Reading proficiency was measured using two instruments, one objective and one subjective. Objective proficiency was assessed with the LexTALE proficiency test (Lemhöfer & Broersma, 2012), a non-speeded visual lexical decision task in which participants classified letter strings as English words or non-words. The percentage of correct responses was used as the objective proficiency score. All other bilingual demographics were measured using selected items from the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007).

Procedure

The procedure was identical across Experiments 1 and 2, with the only difference being the judgement dimension (valence vs arousal): In Experiment 1, participants judged whether stimuli were positive or negative, and in Experiment 2, they judged whether stimuli were calming or arousing. After providing informed consent in German (Curtin University Human Research Ethics Committee Approval number: HREC2020-0026), participants completed three speeded judgement tasks. Participants completed the word judgement task in both their L1 (German) and L2 (English). The order of the language word tasks was counterbalanced. Following this, participants completed a judgement task for the picture stimuli. The picture block was always presented last.

For each trial, a fixation cross was presented for 500 ms, followed by presentation of the stimulus (a word or picture), which remained on screen until response. Participants responded using two keys (E and I). Response labels were presented at the top of the screen and the mapping of labels to keys was counterbalanced by varying whether the target label appeared on the left or on the right; practice trials were presented at the start of the experiment and after label-side switches.

In Experiment 1, participants made valence judgements (positive vs negative; positiv vs. negativ in the German block). In Experiment 2, participants made arousal judgements (exciting vs. calming; aufregend vs ruhig). The materials were identical across experiments. For the word stimuli, 40 words were presented four times each in a random order (160 trials per language block). For the picture stimuli, 48 pictures were presented twice each (96 trials). After completing the judgement tasks, participants completed LexTALE, the English word comprehension check, and the LEAP-Q items.

Analysis

Responses faster than 200 ms and slower than 2500 ms were excluded. Responses inconsistent with the norm-based category were treated as categorisation errors (e.g., labelling a positive item as negative and vice versa, or labelling a high arousal item as low arousal and vice versa) and response times were set to missing. Categorisation errors accounted for 13% of trials in Experiment 1 and 24% in Experiment 2. Remaining response times were screened for outliers (3 standard deviations above or below the mean) and replaced with missing values. For L2 analyses, trials were additionally removed where participants indicated they did not understand the English word in the comprehension check.

Response times were analysed using linear mixed-effects models with participants and items as crossed random effects. Fixed effects included stimulus type (L1 words, L2 words, pictures), valence, arousal, and their interactions, with age and education included as covariates. Stimulus type was dummy coded to compare L2 words with L1 words and pictures; valence and arousal were contrast coded. Models included random intercepts for participants and items, and a by-participant random slope for stimulus type, as follows: RT ~ Valence × Arousal × Stimulus Type + age + education + (1 + Stimulus Type | participant) + (1 | item).

Accuracy was analysed with logistic mixed-effects models using the same fixed-effect structure as the RT models.

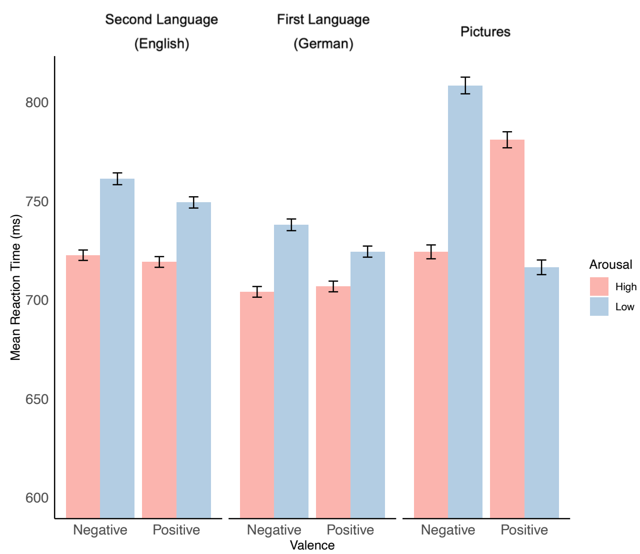

Figure 2: Mean response times across all three stimulus types, valence, and arousal categories for the Valence Judgement Task (Experiment 1). Vertical bars represent errors of the means.

Results

Experiment 1: Valence Judgement Task

Stimulus Type Analysing response times, the mixed-effects model showed a significant valence × arousal interaction (b = -65.47, $p < .001$). Follow-up comparisons indicated that the interaction differed across stimulus types. The conflict pattern was evident for pictures, such that high-arousal negative pictures were responded to faster than low-arousal negative pictures, and low-arousal positive pictures were responded to faster than high-arousal positive pictures (see Figure 2). In contrast, the conflict pattern was not evident for word stimuli in either language. Consistent with this, there was a significant valence × arousal × stimulus type interaction when contrasting L2 words with pictures (b = -145.32, $p < .001$), but no evidence that the interaction differed between L1 and L2 word stimuli (b = 0.55, p = .989), suggesting that words were processed the same across both languages.

Accuracy did not show a reliable overall valence × arousal interaction, b = -0.27, p = .581. However, the accuracy pattern differed between L2 words and pictures, b = 3.66, $p <$ .001. This suggests that the response time pattern was not explained by a simple speed-accuracy trade-off.

Second Language Processing Analyses focusing on L2 words showed that low arousal words were processed slower (b = 37.93, p = .005), however there was no difference in processing between positive and negative words (b = -0.08, p = .995). We found two significant interactive effects related to bilingual demographics. L2 age of acquisition interacted with arousal (b = 2.39, p = .018), such that earlier L2 acquisition was associated with faster responses to low-arousal words, while high-arousal words were relatively less sensitive to age of acquisition.

Given that word frequency has been shown to interact with valence and arousal (e.g., Imbault et al., 2022; Kuperman et al., 2014), we also investigated interactive effects related to frequency. Our findings show that frequency interacted with arousal (b = -88.41, p = .022), whereby low arousal words were processed faster when they were higher frequency. These differences appear to be driven by low frequency words. Specifically, estimated marginal means indicated that response time varied by only 34.67ms between low (M = 707.74ms) or high frequency words (M =742.41). Whereas for low arousal words response time varied by 81.67 between low (M = 803.99ms) and high frequency words (M = 722.32ms).

Experiment 2: Arousal Judgement Task

Stimulus Type Analysing response times, the mixed-effects model showed a significant valence × arousal interaction (b = -114.81, $p < .001$). As shown in Figure 3, the conflict pattern was evident across stimulus types, such that responses were faster for high (vs. low) arousal negative stimuli and for low (vs. high) arousal positive stimuli, consistent with the Valence-Arousal Conflict Pattern (Citron et al., 2014a; Citron et al., 2014b; Purkis et al., 2009; Robinson et al., 2004; Wang et al., 2018). Follow-up comparisons indicated there was a significant valence × arousal × stimulus type interaction when contrasting L2 words with pictures (b = -81.79, p = .039). However, there was no evidence that the interaction differed between L1 and L2 word stimuli (b = 0.55, p = .662), suggesting that words were processed similarly across both languages.

Accuracy showed a valence × arousal interaction, b = 3.82, $p < .001$. This interaction differed between L2 words and pictures, b = -2.26, $p < .001$, but not between L1 and L2 words, b = -0.84, p = .178.

Second Language Processing Analyses focusing on L2 words once again showed a main effect of arousal, whereby low-arousal words were processed more slowly (b = 35.31, p = .004), but no direct effect of valence. We found several significant interactions between bilingual experience variables and affective dimensions. Firstly, daily L2 (English) exposure interacted with arousal (b = -0.59, p = .005), such that higher L2 use was associated with slower responses to high-arousal words, indicating higher language use was linked to increased processing costs specifically for high-arousal words. Secondly, time spent immersed with an L2 family interacted with both valence and arousal. Family immersion interacted with valence (b = -1.40, p = .017), such that greater immersion was associated with slower responses to negative (vs. positive) words. Family immersion also interacted with arousal (b = 1.28, p = .031), such that greater immersion was associated with slower responses to low- (vs. high-) arousal words. These effects indicate that L2 family immersion selectively increased valence- and arousal-based differentiation in RTs rather than shifting overall speed.

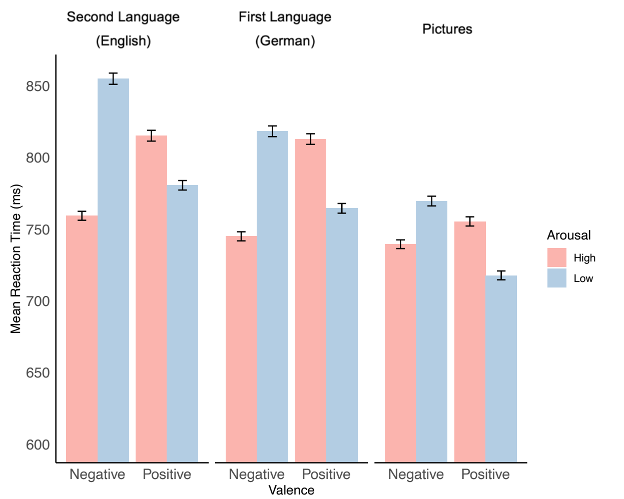

Figure 3: Mean response times across all three stimulus types, valence, and arousal categories for the Arousal Judgement Task (Experiment 2). Vertical bars represent errors of the means.

Discussion

This paper tested whether valence and arousal interact during speeded affective judgements, and whether this interaction depends on stimulus type and language status. Across two experiments, we compared pictures with L1 and L2 word blocks in the same participants, using matched valence × arousal stimulus structure based on normative databases (Bradley & Lang, 1999; Schmidtke et al., 2014; Marchewka et al., 2014).

The main result was a robust valence × arousal interaction, but its observability depended on task demands. In the valence judgement task (Experiment 1), the interaction was present for pictures but not for words. In the arousal judgement task (Experiment 2), the interaction was present for both pictures and words.

These findings contribute in two ways. For dimensional emotion processing accounts, they show that the interaction depends on task demands and is more visible in word stimuli during arousal judgements. For bilingual emotion processing accounts, they show that highly proficient bilinguals did not differ between L1 and L2 in this speeded classification task, although bilingual experience still moderated L2 response times.

Our findings are consistent with the Valence–Arousal Conflict Theory, which predicts faster responses for high-arousal negative and low-arousal positive stimuli, and slower responses for conflicting combinations (Robinson et al., 2004; Purkis et al., 2009; Citron et al., 2014a; Citron et al., 2014b; Wang et al., 2018). It is also consistent with the idea that valence and arousal can interact early in processing, rather than acting as independent dimensions (Citron et al., 2013; Citron et al., 2014a).

While much previous work has investigated emotion effects in terms of valence, our results support the notion that arousal also has a strong effect on semantic processing. This is in line with a meta-analysis of ERP amplitudes which found that high arousal (vs low arousal) stimuli enhanced the negativity bias (within monolingual populations; Yuan et al., 2019), and with a recent meta-analysis suggesting that valence effects in visual word recognition may depend on stimulus characteristics, including the intensity of negative words (Ferré et al., 2025). Taken together, this may suggest that valence-only analyses can be misleading if arousal is not explicitly modelled, as a valence advantage may cancel out when collapsing across high- and low-arousal items. This may contribute to some mixed findings in the broader literature.

Task Demands: Valence vs. Arousal Judgements

The difference between Experiments 1 and 2 suggests that the conflict pattern is easier to observe when the decision requires arousal classification. Arousal judgements may be more difficult than valence judgements (e.g., the beach seems obviously positive in valence, but its arousal is more difficult to determine, as it can be associated with both calming and exciting feelings), suggested by higher error rates and slower processing time (e.g., Purkis et al., 2009). One potential interpretation is that arousal judgements may require more cognitive resources and require access to later, semantic processes. In the case of arousal judgments, participants might need to access and evaluate their internal physiological state. This process might facilitate a deeper semantic integration of the stimuli, making the Valence-Arousal Conflict more apparent. Furthermore, arousal judgments might demand sustained attention over a longer period, which might facilitate the detection of the Valence-Arousal Conflict (Scherer, 2019). Therefore, it is possible that higher task demands increase the contribution of arousal information to response selection, which in turn makes the valence × arousal interaction more visible within word stimuli.

However, it is also possible that the interaction was too small or too noisy to show up in response time data within the faster valence task, because lexical access and other word-level variance may have influenced response times (e.g., frequency; Brysbaert et al., 2018). This account does not assume that the underlying interaction is absent in valence judgements, only that it may be harder to capture within certain experimental paradigms. Indeed, this is consistent with the interactive effect of frequency found in Experiment 1, where the interactive pattern was not found in word stimuli.

Bilingual Affective Processing: L1 vs. L2

Across both experiments, there was no evidence that the interaction differed between L1 and L2 word blocks. This processing pattern suggests that affective information processing may not be restricted to the first (native) language and may appear within an early time window of pre-lexical processing. In line with this view, Herbert (2023) measured event-related potentials (ERPs) in highly proficient German-English bilinguals and found that valence and arousal effects were already observed in an early time window of pre-lexical processing (N100), which precedes lexico-semantic access, suggesting that early emotion appraisals are automatic and not restricted to the L1.

One interpretation is that speeded affective judgements primarily index early evaluative appraisals that are not strongly language-dependent. These appraisals may be available once meaning is accessed, even if later stages of emotion experience or regulation differ across languages. This interpretation is compatible with models in which bilingual word recognition involves parallel activation and early semantic access mechanisms that are not confined to a single language system (Dijkstra et al., 2019).

While previous work has highlighted differences in affective processing (e.g., Kazanas & Altarriba, 2016; Barriga-Paulino et al., 2022), language-dependent emotion differences may not necessarily be confined to early, more automatic semantic processing and may reflect differences in more deliberate stages of emotion, such as conscious emotion regulation (Frijda & Scherer, 2009; Scherer & Moors, 2019). This result is informative for language-dependent emotion accounts, including the foreign language effect, which often predict attenuated affect in L2 (Pavlenko, 2012; Caldwell-Harris, 2014). The present data do not rule out language-dependent effects. Instead, they suggest that language status may not be a primary determinant of early conflict effects in this paradigm.

The Valence-Arousal Conflict Pattern was found within picture stimuli across both experiments. Pictures may trigger affective appraisals more directly than words because they provide richer perceptual input and reduce reliance on lexical access. This aligns with motivated attention accounts in which affective stimuli modulate attention and physiological responding (Lang et al., 1997; Bradley et al., 2001). However, the present data do not show that pictures and words rely on different mechanisms. Instead, they suggest that the interaction may be less reliably expressed in word response times under some task settings, likely because word processing is jointly shaped by affective dimensions and lexical variables (e.g., Ferré et al., 2018; Méndez-Bértolo et al., 2011). Therefore, we suggest that absence of a conflict pattern in word response times does not imply the absence of valence–arousal interaction. It may reflect limited sensitivity under a given task and stimulus set.

Bilingual experience moderated L2 response times in Experiment 2, even though the valence × arousal interaction did not differ between L1 and L2 at the block level. In the L2-only models, higher daily L2 exposure was associated with slower responses to high-arousal words. Greater time spent in an L2 family environment was associated with slower responses to negative (vs. positive) words and to low- (vs. high-) arousal words. These effects did not indicate a general facilitation effect with greater L2 experience. One possibility is that greater experience supports more differentiated semantic evaluation, which can increase response times for some affective categories. Another is that arousal-judgement demands shift cue weighting, and this weighting varies with L2 experience. Such moderation effects are plausible given heterogeneity in bilingual profiles, and they motivate measuring experience in detail rather than treating L2 users as a single group (Grosjean, 2008; De Bruin, 2019; Kaushanskaya et al., 2020).

Limitations

Our work is not without limitations. Because the picture block was always presented after the word blocks, modality effects may partly reflect block order or fatigue. However, this limitation is less relevant for the language comparison because word-block order was counterbalanced.

Furthermore, while we matched our word list on length and frequency, there are still variables which impact word processing which we did not control for. While we attempted to control for as many confounds as possible, it was not possible to include all variables of relevance, therefore this process involved a trade-off, and some variables were not included in the analyses. Variables such as concreteness and emotion-label versus emotion-laden status have been shown to impact processing (e.g., Altarriba, 2006; Yao et al., 2016), which may have impacted our results. This issue is not unique to the current study. For example, Ponari et al. (2015) attempted to control for familiarity, imageability and concreteness in addition to frequency and length, but they had to remove these variables from their analyses as they could not achieve model conversion.

A second limitation is that the present tasks target speeded explicit judgements. Other measures, such as ERPs or pupil dilation, may capture earlier or partially implicit components of affective processing that do not always align with response time patterns (Citron et al., 2013; Toivo & Scheepers, 2019).

Our results provide evidence for a valence–arousal interaction in bilingual affective judgements, but show that its observability depends on task demands and stimulus type. The conflict pattern was robust for pictures in both tasks, and for words in the arousal judgement task. There was no evidence that the interaction differed between L1 and L2 word blocks. These findings support valence–arousal conflict accounts and suggest that early evaluative appraisals, as measured here, may not be strongly language-specific, while bilingual experience can still shape how affective dimensions influence response times.

Acknowledgments

This work has been partially funded by the European Union, by the ERC Advanced Grant Valence Asymmetries (GA n. 101142133). Views and opinions expressed are those of the author only and do not reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Barriga-Paulino, C. I., Guerreiro, M., Faísca, L., & Reis, A. (2022). Does emotional valence modulate word recognition? A behavioral study manipulating frequency and arousal. *Acta Psychologica, 223,* Article e103484.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion, 1*(3), 276.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW). Gainesville, FL: The NIMH Center for the Study of Emotion and Attention, University of Florida.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*(4), 991-997.

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology, 5,* 1055.

Citron, F. M. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014a). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia, 56*(1), 79–89.

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014b). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience, 29,* 1257–1267.

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences, 9,* 1–13.

Dewaele, J.M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics, 40,* 1753–1780.

Dijkstra, T. O. N., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition, 22*(4), 657-679.

Feng, C., Wang, L., Liu, C., Zhu, X., Dai, R., Mai, X., & Luo, Y. J. (2012). The time course of the influence of valence and arousal on the implicit processing of affective pictures. *PloS One,* 7(1), Article e29668.

Ferré, P., Anglada-Tort, M., & Guasch, M. (2018). Processing of emotional words in bilinguals: Testing the effects of word concreteness, task type and language status. *Second Language Research, 34*, 371-394.

Ferré, P., Sánchez-Carmona, A. J., Haro, J., Calvillo-Torres, R., Albert, J., & Hinojosa, J. A. (2025). How does emotional content influence visual word recognition? A meta-analysis of valence effects. *Psychonomic Bulletin & Review, 32*(2), 570-587.

Grosjean, F. (2008). *Studying bilinguals.* Oxford University Press.

Harris, C., Gleason, J. & Ayçiçegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), Bilingual Minds (Emotional experience, expression, and representation) (pp. 257– 283). De Gruyter.

Herbert, C. (2023). Early, emotional and embodied? Processing of emotional words and body words in the native and a second language–evidence from early event-related brain potential modulation and rapid serial visual presentation. *Language, Cognition and Neuroscience, 38*(10), 1384-1411.

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience, 35*(7), 813-839.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience, 9,* 389-397.

Imbault, C., Titone, D., Warriner, A. B., & Kuperman, V. (2022). How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. *Bilingualism: Language and Cognition, 24*(2), 281-292.

Jankowiak, K., & Korpal, P. (2018). On modality effects in bilingual emotional language processing: Evidence from galvanic skin response. *Journal of Psycholinguistic Research, 47,* 663-677.

Jończyk, R. (2016). *Affect-language interactions in native and non-native English speakers: A neuropragmatic perspective.* Springer International Publishing AG.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition, 23,* 945-950.

Kazanas, S. A., & Altarriba, J. (2016). Emotion word processing: Effects of word type and valence in Spanish–English bilinguals. *Journal of Psycholinguistic Research, 45,* 395- 406.

Lang, P. J. (1980). Behavioral treatment and biobehavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), Technology in mental health care delivery systems (pp. 119-l 37). Norwood, NJ: Ablex.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist, 50*(5), 372.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), Attention and orienting: Sensory and motivational processes (pp. 97–135). Hillsdale, NJ: Erlbaum.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44,* 325-343.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25,* 605- 621.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*(2), 596-610.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940–967.

Méndez-Bértolo, C., Pozo, M. A., & Hinojosa, J. A. (2011). Word frequency modulates the processing of emotional words: Convergent behavioral and electrophysiological data. *Neuroscience Letters, 494*(3), 250-254.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47,* 405–428.

Purkis, H. M., Lipp, O. V., Edwards, M. S., & Barnes, R. (2009). An increase in stimulus arousal has differential effects on the processing speed of pleasant and unpleasant stimuli. *Motivation and Emotion, 33,* 353–361.

Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! that could be dangerous: Valence-arousal interactions in evaluative processing. Personality & *Social Psychology Bulletin, 30,* 1472–1484.

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods, 46,* 1108-1118.

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology, 70,* 719-745.

Stankovic, M., Hamamura, T., & Biedermann, B. (2024). German-English stimulus set: Affective and psycholinguistic matching [Preprint and Data Set]. OSF. https://doi.org/10.17605/OSF.IO/Y6PXF

Wang, Y. M., Li, T., & Li, L. (2018). Valence evaluation with approaching or withdrawing cues: Directly testing valence–arousal conflict theory. *Cognition & Emotion, 32*(4), 904-912.

Yuan, J., Tian, Y., Huang, X., Fan, H., & Wei, X. (2019). Emotional bias varies with stimulus type, arousal and task setting: Meta-analytic evidences. *Neuroscience & Biobehavioral Reviews, 107*, 461-472.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions In Psychological Science, 10*(6), 224-228.